**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**

[insert text here]

**Abstract**

[insert text here]

**Acknowledgements**

I thank my sanity for holding together just long enough…

Contents

[List of Tables 10](#_Toc161657168)

[List of Appendix Tables 12](#_Toc161657169)

[List of Figures 13](#_Toc161657170)

[List of Abbreviations 14](#_Toc161657171)

[Part 1 Introduction 15](#_Toc161657172)

[Chapter 1.1 Social Theory 17](#_Toc161657173)

[Chapter 1.2 School-to-work transitions in context 22](#_Toc161657174)

[Chapter 1.3 Social Stratification 24](#_Toc161657175)

[Chapter 1.4 The British Education System 24](#_Toc161657176)

[Chapter 1.5 Research Questions 27](#_Toc161657177)

[Chapter 1.6 Data and Methods 28](#_Toc161657178)

[1.6.1 National Childhood Development Study 29](#_Toc161657179)

[1.6.2 British Cohort Survey 31](#_Toc161657180)

[1.6.3 The British Household Panel Survey 32](#_Toc161657181)

[1.6.4 United Kingdom Household Panel Survey 34](#_Toc161657182)

[1.6.4.1 Complex Survey Design 35](#_Toc161657183)

[1.6.5 Next Steps 36](#_Toc161657184)

[Chapter 1.7 Methods 37](#_Toc161657185)

[Chapter 1.8 Structure of Thesis 44](#_Toc161657186)

[Part 2 The National Childhood Development Survey (1958): Youth Transitions in Education and Employment 45](#_Toc161657187)

[Chapter 2.1 Introduction 45](#_Toc161657188)

[Chapter 2.2 Literature Review: NCDS Timeframe and Context 47](#_Toc161657189)

[2.2.1 Story of transitions for NCDS youth 48](#_Toc161657190)

[Structural Barriers to successful transitions – the role of sex and social-class 57](#_Toc161657191)

[2.2.1.1 Sex 57](#_Toc161657192)

[2.2.1.2 Social Class 59](#_Toc161657193)

[2.2.1.3 Educational Attainment and training 59](#_Toc161657194)

[Chapter 2.3 Data and Methods 61](#_Toc161657195)

[2.3.1 Introduction to measures for subsequent analysis 63](#_Toc161657196)

[2.3.1.1 Economic Activity 64](#_Toc161657197)

[2.3.1.2 Educational Attainment 68](#_Toc161657198)

[2.3.1.3 Sex 71](#_Toc161657199)

[2.3.1.4 Race 71](#_Toc161657200)

[2.3.1.5 Housing Tenure 72](#_Toc161657201)

[2.3.1.6 Social Stratification: NS-SEC, CAMSIS, RGSC 72](#_Toc161657202)

[2.3.1.6.1 SOC Codes and the construction of socio-economic variables 75](#_Toc161657203)

[2.3.1.6.2 Registrar General Class Schema 82](#_Toc161657204)

[2.3.1.6.3 National Statistics Socio-Economic Classification 85](#_Toc161657205)

[2.3.1.6.4 CAMSIS 88](#_Toc161657206)

[Chapter 2.4 Descriptive Statistics 91](#_Toc161657207)

[Chapter 2.5 Modelling Main Economic Activity 108](#_Toc161657208)

[2.5.1 Discussion and Conclusion 129](#_Toc161657209)

[Chapter 2.6 Sensitivity Analysis of Independent Variables 132](#_Toc161657210)

[2.6.1 Testing Measures of Parental Social Class 133](#_Toc161657211)

[2.6.2 Discussion and Conclusions 141](#_Toc161657212)

[Chapter 2.7 Sensitivity analysis using SOC codes 142](#_Toc161657213)

[2.7.1 SOC codes Modelling 142](#_Toc161657214)

[2.7.2 Discussion and Conclusion 155](#_Toc161657215)

[Chapter 2.8 Missing Data in the NCDS 157](#_Toc161657216)

[2.8.1 Missing Data 157](#_Toc161657217)

[2.8.2 Simulation Study 160](#_Toc161657218)

[2.8.3 Multiple Imputation by Chained Equations 172](#_Toc161657219)

[2.8.4 Discussion and Conclusions 187](#_Toc161657220)

[Chapter 2.9 Discussion and Conclusions for Part 1 188](#_Toc161657221)

[Part 3 The British Cohort Study: Youth Transitions in Education and Employment 194](#_Toc161657222)

[Chapter 3.1 Introduction 194](#_Toc161657223)

[Chapter 3.2 Literature Review: BCS Timeframe and Context 195](#_Toc161657224)

[3.2.1 Story of transitions for BCS youth 196](#_Toc161657225)

[3.2.2 Structural Barriers to successful transitions – the role of social class and sex 202](#_Toc161657226)

[3.2.2.1 Social Class 202](#_Toc161657227)

[3.2.2.2 Sex 203](#_Toc161657228)

[3.2.2.3 Conclusion 203](#_Toc161657229)

[Chapter 3.3 Data and Methods 205](#_Toc161657230)

[3.3.1 Introduction to measures for subsequent analysis 209](#_Toc161657231)

[3.3.1.1 Economic Activity 209](#_Toc161657232)

[3.3.1.2 Educational Attainment 212](#_Toc161657233)

[3.3.1.3 Sex 213](#_Toc161657234)

[3.3.1.4 Housing Tenure 214](#_Toc161657235)

[3.3.1.5 Social Stratification and Socio-Economic Background: NS-SEC, CAMSIS, RGSC 215](#_Toc161657236)

[3.3.1.5.1 SOC Codes 216](#_Toc161657237)

[Chapter 3.4 Descriptive Statistics 217](#_Toc161657238)

[Chapter 3.5 Modelling Main Economic Activity: 231](#_Toc161657239)

[3.5.1 Discussion and Conclusion 244](#_Toc161657240)

[Chapter 3.6 Sensitivity Analysis of Independent Variables 246](#_Toc161657241)

[3.6.1 Testing Measures of Parental Social Class 247](#_Toc161657242)

[3.6.2 Discussion and Conclusions 255](#_Toc161657243)

[Chapter 3.7 Sensitivity analysis using SOC codes 256](#_Toc161657244)

[3.7.1 SOC Codes Modelling 256](#_Toc161657245)

[3.7.2 Discussion and Conclusion 261](#_Toc161657246)

[Chapter 3.8 Missing Data in the BCS 262](#_Toc161657247)

[3.8.1 Discussion and Conclusions 276](#_Toc161657248)

[Chapter 3.9 Discussion and Conclusions for Part 2 278](#_Toc161657249)

[Part 4 The United Kingdom Household Panel Survey 281](#_Toc161657250)

[Chapter 4.1 Introduction to Part 3 281](#_Toc161657251)

[Chapter 4.2 Literature Review: UKHLS Timeframe and Context 281](#_Toc161657252)

[4.2.1 Story of transitions for UKHLS Youth 281](#_Toc161657253)

[4.2.2 Structural barriers to successful transitions – the role of social class and sex 281](#_Toc161657254)

[Chapter 4.3 Data and Methods 281](#_Toc161657255)

[4.3.1 Introduction to the UKHLS data 281](#_Toc161657256)

[4.3.2 Synthetic Cohorts 281](#_Toc161657257)

[4.3.3 Introduction to measures for subsequent analysis 281](#_Toc161657258)

[Chapter 4.4 Descriptive Statistics 281](#_Toc161657259)

[Chapter 4.5 Modelling Main Economic Activity 281](#_Toc161657260)

[4.5.1 Discussion and Conclusions 281](#_Toc161657261)

[Chapter 4.6 Sensitivity Analysis of Independent Variables 281](#_Toc161657262)

[4.6.1 Testing Measures of Parental Social Class 281](#_Toc161657263)

[4.6.2 Discussion and Conclusions 281](#_Toc161657264)

[Chapter 4.7 Missing Data in the UKHLS 281](#_Toc161657265)

[4.7.1 Discussion and Conclusions 281](#_Toc161657266)

[Chapter 4.8 Duplication Analysis using Next Steps 281](#_Toc161657267)

[4.8.1 Data and Methods 281](#_Toc161657268)

[4.8.1.1 Introduction to Next Steps Data 281](#_Toc161657269)

[4.8.1.2 Introduction to measures for subsequent analysis 281](#_Toc161657270)

[4.8.2 Descriptive Statistics 281](#_Toc161657271)

[4.8.3 Modelling Main Economic Activity 281](#_Toc161657272)

[4.8.3.1 Discussion and Conclusions 281](#_Toc161657273)

[4.8.4 Sensitivity Analysis 281](#_Toc161657274)

[4.8.4.1 Discussion and Conclusions 282](#_Toc161657275)

[4.8.5 Missing Data in Next Steps 282](#_Toc161657276)

[4.8.5.1 Discussion and Conclusions 282](#_Toc161657277)

[Chapter 4.9 Discussion and Conclusions for Part 3 282](#_Toc161657278)

[Part 5 Comparison of NCDS, BCS, UKHLS, and Next Steps Cohorts 283](#_Toc161657279)

[Chapter 5.1 Introduction to Part 4 283](#_Toc161657280)

[Chapter 5.2 The effects of structural inequality across cohorts 283](#_Toc161657281)

[5.2.1 Discussion and Conclusions 283](#_Toc161657282)

[Part 6 Conclusions 283](#_Toc161657283)

[Chapter 6.1 Introduction to Part 5 283](#_Toc161657284)

[Chapter 6.2 Substantive Conclusions 283](#_Toc161657285)

[Chapter 6.3 Methodological Reflections 283](#_Toc161657286)

[Chapter 6.4 Final Remarks 283](#_Toc161657287)

[Part 7 Appendix: 284](#_Toc161657288)

[Chapter 7.1 Appendix One: NCDS 284](#_Toc161657289)

[Chapter 7.2 Appendix Two: BCS 295](#_Toc161657290)

[Part 8 Bibliography 305](#_Toc161657291)

###### of Tables

[Table 1.1 Sweeps Included in Analysis NCDS 30](#_Toc161657292)

[Table 1.2 Sweeps Included for Analysis BCS 31](#_Toc161657293)

[Table 1.3 Sweeps included in Analysis Next Steps 36](#_Toc161657294)

[Table 2.1 Participation in the NCDS from birth to 23 years 62](#_Toc161657295)

[Table 2.2 Frequency Statistics for Economic Activity 65](#_Toc161657296)

[Table 2.3 Educational Attainment Count Variable by Economic Activity 70](#_Toc161657297)

[Table 2.4 Breakdown of classification of SOC 90 and SOC 2000 78](#_Toc161657298)

[Table 2.5 Sub-major groups of SOC 90 and SOC 2000 by Skill Level 78](#_Toc161657299)

[Table 2.6 RGSC Class Schema 84](#_Toc161657300)

[Table 2.7 NS-SEC Class Schema 87](#_Toc161657301)

[Table 2.8 Examples of Occupations from Analytical NS-SEC 88](#_Toc161657302)

[Table 2.9 Examples of CAMSIS scores by SOC-90 Codes 90](#_Toc161657303)

[Table 2.10 Descriptive Statistics for Economic Activity Model 92](#_Toc161657304)

[Table 2.11 Descriptive Statistics by Economic Activity 95](#_Toc161657305)

[Table 2.12 Descriptive Statistics comparing NS-SEC by SOC2000 and SOC90 codes 105](#_Toc161657306)

[Table 2.13 Descriptive Statistics comparing RGSC by SOC2000 and SOC90 codes 107](#_Toc161657307)

[Table 2.14 Descriptive Statistics comparing CAMSIS by SOC2000 and SOC90 codes 108](#_Toc161657308)

[Table 2.15 Goodness-of-fit summaries for explanatory variables and Economic Activity 109](#_Toc161657309)

[Table 2.16 Model building goodness-of-fit summaries for multinominal logistic regression model of Economic Activity 110](#_Toc161657310)

[Table 2.17 Mlogit of Economic Activity 118](#_Toc161657311)

[Table 2.18 Sensitivity analyses of alternative measures of parental social stratification 137](#_Toc161657312)

[Table 2.19 Goodness-of-fit summaries for explanatory variables and Economic Activity Comparing SOC codes 143](#_Toc161657313)

[Table 2.20 Model building goodness-of-fit summaries for multinominal logistic regression model of Economic Activity Comparing SOC codes 144](#_Toc161657314)

[Table 2.21 Sensitivity analyses of SOC codes 152](#_Toc161657315)

[Table 2.22 Missing data patterns for NCDS 171](#_Toc161657316)

[Table 2.23 Comparison of CRA NS-SEC vs Imputed NS-SEC 182](#_Toc161657317)

[Table 3.1 Participation in the BCS from Birth to 30 years 208](#_Toc161657318)

[Table 3.2 Frequency Statistics for Economic Activity 211](#_Toc161657319)

[Table 3.3 Descriptive Statistics for Economic Activity Model 220](#_Toc161657320)

[Table 3.4 Descriptive Statistics by Economic Activity 224](#_Toc161657321)

[Table 3.5 Descriptive Statistics Comparing NS-SEC by SOC2000 and SOC90 Codes 228](#_Toc161657322)

[Table 3.6 Descriptive Statistics comparing RGSC by SOC2000 and SOC90 codes 230](#_Toc161657323)

[Table 3.7 Descriptive Statistics comparing CAMSIS by SOC2000 and SOC90 codes 231](#_Toc161657324)

[Table 3.8 Goodness-of-fit summaries for explanatory variables and Economic Activity 232](#_Toc161657325)

[Table 3.9 Model building goodness-of-fit summaries for multinominal logistic regression model of Economic Activity 232](#_Toc161657326)

[Table 3.10 Mlogit of Economic Activity 236](#_Toc161657327)

[Table 3.11 Sensitivity analyses of alternative measures of parental social stratification 251](#_Toc161657328)

[Table 3.12 Goodness-of-fit summaries for explanatory variables and Economic Activity Comparing SOC codes 256](#_Toc161657329)

[Table 3.13 Model building goodness-of-fit summaries for multinominal logistic regression model of Economic Activity Comparing SOC codes 257](#_Toc161657330)

[Table 3.14 Sensitivity Analysis of SOC Codes 258](#_Toc161657331)

[Table 3.15 Missing data patterns for BCS 263](#_Toc161657332)

[Table 3.16 Number of Observations missing for BCS 263](#_Toc161657333)

[Table 3.17 Comparison of Missingness across four models 273](#_Toc161657334)

###### of Appendix Tables

[Appendix 6.1.1 Goodness-of-fit summaries for explanatory variables and Economic Activity (CAMSIS) 284](#_Toc161657335)

[Appendix 6.1.2 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (CAMSIS) 284](#_Toc161657336)

[Appendix 6.1.3 Goodness-of-fit summaries for explanatory variables and Economic Activity (RGSC) 285](#_Toc161657337)

[Appendix 6.1.4 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (RGSC) 285](#_Toc161657338)

[Appendix 6.1.5 Goodness-of-fit summaries for explanatory variables and Economic Activity (CAMSIS) using SOC2000 and SOC90 Codes 286](#_Toc161657339)

[Appendix 6.1.6 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (CAMSIS) using SOC2000 and SOC90 Codes 286](#_Toc161657340)

[Appendix 6.1.7 Goodness-of-fit summaries for explanatory variables and Economic Activity (RGSC) using SOC2000 and SOC90 Codes 287](#_Toc161657341)

[Appendix 6.1.8 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (RGSC) using SOC200 and SOC90 Codes 287](#_Toc161657342)

[Appendix 6.1.9 Sensitivity Analysis of SOC Codes (CAMSIS) 289](#_Toc161657343)

[Appendix 6.1.10 Sensitivity Analysis of SOC Codes (RGSC) 292](#_Toc161657344)

[Appendix 6.2.1 Goodness-of-fit summaries for explanatory variables and Economic Activity (CAMSIS) 295](#_Toc161657345)

[Appendix 6.2.2 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (CAMSIS) 295](#_Toc161657346)

[Appendix 6.2.3 Goodness-of-fit summaries for explanatory variables and Economic Activity (RGSC) 296](#_Toc161657347)

[Appendix 6.2.4 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (RGSC) 296](#_Toc161657348)

[Appendix 6.2.5 Goodness-of-fit summaries for explanatory variables and Economic Activity (CAMSIS) using SOC2000 and SOC90 Codes 297](#_Toc161657349)

[Appendix 6.2.6 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (CAMSIS) using SOC2000 and SOC90 Codes 297](#_Toc161657350)

[Appendix 6.2.7 Goodness-of-fit summaries for explanatory variables and Economic Activity (NS-SEC) using SOC2000 and SOC90 Codes 298](#_Toc161657351)

[Appendix 6.2.8 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (NS-SEC) using SOC2000 and SOC90 Codes 298](#_Toc161657352)

[Appendix 6.2.9 Sensitivity Analysis of SOC Codes (CAMSIS) 300](#_Toc161657353)

[Appendix 6.2.10 Sensitivity Analysis of SOC Codes (NS-SEC) 302](#_Toc161657354)

###### of Figures

[Figure 1.1 Transitional Pathways for NCDS Cohort 47](#_Toc161657355)

[Figure 1.2 Predicted Probabilities of Economic Activity by NS-SEC 123](#_Toc161657356)

[Figure 1.3 Predicted Probabilities of Economic Activity by Sex 124](#_Toc161657357)

[Figure 1.4 Predicted Probabilities of Economic Activity by Educational Attainment 125](#_Toc161657358)

[Figure 1.5 Predicted Probabilities of Economic Activity by Housing Tenure 126](#_Toc161657359)

[Figure 1.6 Log Odds versus Quasi-variance statistics of individuals being in School over Employment 127](#_Toc161657360)

[Figure 1.7 Log Odds versus Quasi-variance statistics of individuals being in School over Unemployment & OLF 127](#_Toc161657361)

[Figure 1.8 Log Odds versus Quasi-variance statistics of individuals being in School over Training & Apprenticeship 128](#_Toc161657362)

[Figure 1.9 Log Odds versus Quasi-variance statistics of individuals being in School over Non-traditional Education 128](#_Toc161657363)

[Figure 1.10 Trace plot summaries for Economic Activity 179](#_Toc161657364)

[Figure 1.11 Trace plot summaries for Educational Attainment 179](#_Toc161657365)

[Figure 1.12 Trace plot summaries for NS-SEC 180](#_Toc161657366)

[Figure 1.13 Trace plot summaries for Housing Tenure 180](#_Toc161657367)

[Figure 2.1 Predicted Probabilities of Economic Activity by NS-SEC 240](#_Toc161657368)

[Figure 2.2 Predicted Probabilities of Economic Activity by Educational Attainment 240](#_Toc161657369)

[Figure 2.3 Predicted Probabilities of Economic Activity by Sex 241](#_Toc161657370)

[Figure 2.4 Predicted Probabilities of Economic Activity by Housing Tenure 241](#_Toc161657371)

[Figure 2.5 Log Odds versus Quasi-variance statistics of individuals being in Education over Employment 243](#_Toc161657372)

[Figure 2.6 Log Odds versus Quasi-variance statistics of individuals being in Education over Training & Apprenticeships 243](#_Toc161657373)

[Figure 2.7 Log Odds versus Quasi-variance statistics of individuals being in Education over Unemployment & OLF 244](#_Toc161657374)

[Figure 2.8 Trace plot summaries for Economic Activity 266](#_Toc161657375)

[Figure 2.9 Trace plot summaries for Educational Attainment 266](#_Toc161657376)

[Figure 2.10 Trace plot summaries for RGSC 267](#_Toc161657377)

[Figure 2.11 Trace plot summaries for Housing Tenure 267](#_Toc161657378)

###### of Abbreviations

A’level Advanced Level

BCS British Cohort Survey

CRA Complete Records Analysis

CSE Certificate of Secondary Education

ISCO International Standard Classification of Occupations

MI Multiple Imputation

NCDS National Childhood Development Study

NEET Not in Education, Employment, or Training

NS-SEC National Statistics Socio-economic classification

NVQ National Vocational Qualification

O’Level Ordinary Level

RGSC Registrar General’s Social Classes

SOC Standard Occupational Classification

TOPs Training Opportunities Scheme

UKHLS United Kingdom Household Longitudinal Study

GCSE General Certificate of Secondary Education

ONS Office for National Statistics

ERSC Economic Social Research Council

NDYP New Deal for Young People

# Introduction

There is a strong research tradition studying the role that structural inequalities play in youth transitions (Furlong and Cartmel, 1997; Bynner, 1998, 1999; Croxford *et al.*, 2006; Duckworth and Schoon, 2012; Dorsett and Lucchino, 2013; Duta and Iannelli, 2018; Duta, Wielgoszewska and Iannelli, 2020). Empirical research has consistently demonstrated that structural inequalities such as social class, sex, and housing tenure all 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 the research tradition of the sociology of youth transitions. First, it provides new empirical evidence about the nature of structural inequalities influence within and between cohorts, for the first time creating synthetic cohorts to study and understand the overlooked period of the 1990s. 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 the area of youth transitions by handling missing data, accomplished by discussing and implementing different handling missing data techniques. All three contributions seek to understand structural inequalities influence upon youth’s first large decision in life, the school-to-work transition following mandatory education and how these structural inequalities may have changed and evolved as time has progressed.

There are four concepts that will be studied in this thesis that come under the umbrella of ‘structural inequalities’: social class, sex, housing tenure, and prior educational attainment. Social class is a broad concept that has many manifestations. Due to this, as mentioned, a sensitivity analysis of measures will be conducted using different theoretical and empirical constructions used to understand if use of one measure over another alters the substantive conclusions reached in the study of youth transitions. Sex is the second structural inequality that is included in the thesis. Whilst there is an argument of whether the inclusion of Sex or Gender is more appropriate, the inclusion of older datasets where only Sex was recorded makes this a relatively simple decision to make. Housing tenure is the third structural inequality to be included in the analysis. Housing tenure is defined as the type of household that an individual grew up in – a home owned by their parents, or a home not owned by their parents. Educational attainment is the final measure included within the thesis for analysis. For the older cohorts, educational attainment is measured as a child’s school-level Ordinary level or O’level attainment. O’levels are a set of examinations historically taken at the end of compulsory schooling in England and Wales[[1]](#footnote-1). For the younger cohorts included in this analysis O’levels were replaced with the General Certificate of Secondary Education (GCSE) in 1986. 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 all presented, and the main overall structure of the thesis is outlined.

## Social Theory

The life course is a term that will be consistently used within this thesis. It is a fundamental aspect of subsequent analysis to understand how youth transitions can be more comprehensively understood within a life course perspective. The life course dispenses with static ‘snapshot’ notions of sociology. Instead, it views the individual in a constant web of changing temporal contexts influencing the agent (Elder, 1994). The life course approach is best suited for analysing youth transitions using longitudinal data. It incorporates the changing processes and influences that ultimately impact an individual’s choices and opportunities when engaging in transitions during the youth stage. The use of repeated contacts-based data that is used in this thesis, thus makes the life course approach an attractive form of social theory to engage with.

The life course approach has established itself as a substantively significant research paradigm within the last few decades (Elder, 1994). The term ‘life course’ is a concrete multilevel phenomenon defined via individuals' social trajectories through structured pathways of given institutions that form the developmental experience of a given individual (Elder, 1994). These ‘structured pathways’ are interwoven with what Elder argued were ‘age-graded trajectories’ (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 into at specific points in their lives that are ultimately influenced and dependent upon structural inequalities. This is almost verbatim a description of the life course explained through Elder (1994).

This thesis will study the impact of structural inequalities of an individual’s ‘choice’ and opportunity post-mandatory schooling. The word choice is deliberately encapsulated in quotes to emphasise the somewhat controversial nature of the term. Choice is dependent upon many aspects of an individual’s life. 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, for the most part, already been made, the parents cannot afford to send the child to private school and so the child will not go. In this respect, ‘choice’ is little more than an illusion for the working-class family. Whilst there are avenues for a working-class child to attend private school – 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 asterix, 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). In other areas of sociology this is also termed ‘bounded agency’ (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).

The outcome of a child's post-mandatory schooling impacts their life chances across their 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 sociological tools 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 on school-to-work transitions. Structural inequalities are viewed through the lens of the role of structure. This role of structure is primarily expressed through the social theory of structuration (Giddens, 1989). Structuration argues that structural factors like social class, gender, and ethnicity 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). A retort to the theory of Structuration comes in the form of the theory of individualisation. The latter 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).

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 ‘detraditionalisation’ (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). In other words, individualisation fails to account for structural elements' still apparently strong influence on a 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.

Proponents of structuration, such as Giddens (1989), appear not to appreciate the increasing complexity levels placed upon individuals, leading to potentially heterogenous outcomes even if the effect of structural inequalities remains strong. In this case, it is best to call for a ‘Structured Individualism’ thesis. 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 upon power relations (Hitlin and Johnson, 2015; Schmitt, 2021).

Whilst Individualisation has proved very difficult to establish in the empirical sociology cannon (Gayle, Lambert and Murray, 2009), Structuration as a theory has not been immune to robust critique from other sources of social theory. A belief that social class is declining in relevance amongst British society is central to the argument of ‘New Structuralism’ which grew out of the 1980s. The argument follows that new structural cleavages have arisen over and above class-based effects (Devine, 2017). This New Structuralism argues that consumption-based cleavages – most notably related to housing tenure – are more influential on outcomes than social class. Whilst there is existing evidence that persistent class effects on the outcomes of young people are present (Micklewright, 1989). These older studies have not been updated with contemporary statistical techniques and should be re-analysed.

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, Structuration, individualisation, New Structuralism, and Structured Individualism all offer practical explanations that often overlap and compete with one another. 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.

## School-to-work transitions in context

Youth transitions cover two of three life domains defined by Mayer and Schoepflin (2022). Specifically, the life domain of education and preparation for work and the phase of active employment overlap with a subdivision of youth transition studies known as school-to-work transitions. School-to-work transitions have a rich sociological tradition within empirical youth transitions research (Clarke, 1978; Raffe, 1984; Bynner, 1998, 1999; Gayle, 1998; Vickerstaff, 2003; Croxford *et al.*, 2006; Brooks, 2009; Iannelli and Smyth, 2017). There is general agreement that the socio-historical context that young people grew up in within the latter half of the 20th century has dramatically transformed from the opening decades of the century (Murray and Gayle, 2012). Since the end of the second world war, young people have remained in education beyond the compulsory period, the youth labour market collapsed in the 1980s, apprenticeship schemes were removed and replaced with youth training schemes, and Britain went through vast economic re-structuring moving from an industrial, to post-industrial economy (ibid). All of these factors are bound together in what Gayle, Lambert, and Murray (2009) have labelled the ‘changing times consensus’. Change is the defining factor of school-to-work transitions in the latter half of the 20th century, this change is an amalgamation of exogenous economic shocks like economic restructuring and recessions, and state acted social policy such as the establishment of modern apprenticeships, the raising of the school leaving age, and the New Deal for Young People (NDYP).

Previous research on British school-to-work transitions has focused on a comparison of change between cohorts from different time points (Bynner, 1998, 1999; Bynner and Ferri, 2003). This research has mainly 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 timepoints. Unfortunately, no birth cohorts exist from 1970 to the end of the 20th century, presenting a large gap in the study of school-to-work transitions and the overall investigation of change. Contemporary statistical approaches provide the ability to construct synthetic cohorts using non-birth cohort data to finally study this gap.

## 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 creates 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 and 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).

## The British Education System

The start of the latter half of the 20th century has been described as ‘capitalism’s golden age’ (Hobsbawn in Jones, 2016). 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 – eventually 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 1965 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, this statistic was over 90 per cent for Scottish school leavers that left at age 15 (ibid). The re-structuring of the British economy saw an increased demand for certification and credentialisation (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 that of the General Certification of Education was introduced in 1963 (with 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.

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.

## 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 and sometimes complex social science survey data.

The thesis itself is split into three distinct though inter-related and connected parts. In Part 1, a historical study of structural inequalities is undertaken by looking at the youth of the National Childhood Development Study. 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 British Cohort Study – a cohort of people that were born 12 years after the National Childhood Development Study. The analysis in part 2 will attempt to duplicate the analysis conducted in part 1, firstly to provide an answer to the first research question in the context of a new cohort but also to begin to answer three further research questions:

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

The same pattern will exist for part 3 of the thesis, whereby synthetic cohorts of the United Kingdom Household Panel Survey will be constructed in order to duplicate the analysis conducted in parts 1 and 2 of the thesis. Parts 1-3 of the thesis will primarily concern themselves with research question one. A fourth part to the thesis will act as a comparison section that seeks to bring together the results found across parts 1-3 to answer questions 2-4.

## Data and Methods

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. The first birth cohort study is the National Childhood Development Study (NCDS). The second birth cohort study is the British Cohort Study (BCS). The first household panel survey is the British Household Panel Survey (BHPS). The second household panel survey is the United Kingdom Household Panel Survey (UKHLS). 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. The following sections outline the relevance of each database for inclusion in this thesis, the issues and considerations that result with 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 (University College London, UCL Institute of Education, and Centre for Longitudinal Studies, 2023)[[2]](#footnote-2). The NCDS is a nationally representative birth cohort study that follows the lives of individuals from England, Wales, and Scotland from birth (Power and Elliott, 2006). The NCDS follows 17,415 participants using a cross-sectional sampling design to collect participants from birth within the week of 3-9 of 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 life as they entered adulthood. The NCDS managed to obtain information on 98 per cent of total births during the week between the 3rd and 9th of March 1958 (Shepherd, 1995). 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 of the cohort members school, and finally a test booklet completed by the cohort member in school. The second and third follow-up was 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. 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. This sweep consisted of a cohort member interview that was undertaken by a market research interviewer and supplemented by the 1971 and 1981 census. 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 greater majority of those involved in the NCDS (Shepherd, 1995). The sampling strategy for the fourth follow-up differed from prior sweeps. The sampling strategy only included individuals that has 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 1.1 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 |

### British Cohort Survey

This thesis will also use data from the British Cohort Study (University College London, 2022). The BCS70 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. 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 subsample taken at 21 (Bynner, 2017). Initially, the data was collected using medical records and the mother’s input. As the cohort members got older, 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).

Table 1.2 Sweeps Included for Analysis BCS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | 1970 | 1980 | 1991 | 1996 | 2000 |
| Sweep | 0 | 2 | (sub-sample) | 4 | 5 |
| Sample Size [Include when full re-write of BCS is complete] |  |  |  |  |  |
| Age | Birth | 10 | 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 datasets 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 virtually identical. The 21 subsample, however, provides 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 British Household Panel Survey

The British Household Panel Survey (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 original households, all adult members of their new households would be interviewed. Children 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 8167 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 household into a house with one or more new people, or one or more people move 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 provides 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 also collected educational data for youth once they entered the adult survey for the first time.

### United Kingdom Household Panel Survey

The United Kingdom Household Panel Survey (UKHLS) builds upon the BHPS, incorporating all members into the UKHLS study. The UKHLS started in January of 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). 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 form the ethnic minority boost, 1,500 from the Innovation Panel, and 8,400 from the BHPS participants at Wave 18 of study. The total achieved were 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.

#### Complex Survey Design

Both the BHPS and the UKHLS are complex survey designs. The sampling strategies that both employed makes a straightforward analysis ill-advised. It is vital to reflect on the design and sampling strategies of both household panel surveys to make appropriate adjustments to the subsequent analysis. 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 and the ethnic minority boost sample.

Survey weights are provided by both the BHPS and the UKHLS to handle non-response and deal with the complex survey design of the household panel surveys. Advice on what weights to use and why is provided by Lynn and Kaminska (2010). On top of this complex survey design packages within stata such as the ‘svy’ package make suitable adjustments for complex designs and sampling.

### Next Steps

Next Steps, formerly known as the Longitudinal Study of Young People in England (LSYPE) is a longitudinal study that follows a sample of around 16,000 people born in 1989/90. Study members were originally recruited through schools in England when they were 13-14 in 2004. They have been interviewed annually since then. Parents that lived with the children were also interviewed within the first four waves (years) of the study. Further educational data was provided through the linkage of the NPD. Data collection involved in-person interviews for the first four waves of study but from wave five onwards a sequential mixed-mode design is used (Calderwood and Sanchez, 2016). If participants did not respond to initial query, a telephone questionnaire was set up to tackle non-response. Next Steps uses a two-stage probability proportional to size (PPS) sampling procedure with disproportionate stratification. The primary sampling units were schools and the sampling frame includes all maintained and independent schools as well as Pupil Referral Units in England during February 2014. The second stage of sampling selected pupils using the Pupil Level Annual School Census for state schools with probability of selection depending on ethnic groups. For pupils in independent schools and pupil referral units they were sample directly from their school rolls (ibid).

Table 1.3 Sweeps included in Analysis Next Steps

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | 2004 | 2005 | 2006 | 2007 | 2008 |
| Sweep | 0 | 1 | 2 | 3 | 4 |
| Sample Size [Include Here] |  |  |  |  |  |
| Age | 14 | 15 | 16 | 17 | 18 |

## Methods

All research questions in this thesis are answered using quantitative methods research. The subsequent work uses large-scale, complex datasets. The following work on each cohort is broken down into three substantive sections covering: an initial multinominal logistic regression, followed by a sensitivity analysis of social stratification measures and Standard Occupation Classification codes, followed by a handling missing data section. Each model provided will contain log odds, average marginal effects, and quasi-variance statistics 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 analysis. All analyses share a common use of multinominal logistic regression models and as such these models will be outlined now. The multinominal logistic regression model is a non-parametric, non-linear model that is an extension of the logistic regression model that handles a nominal categorical dependent variable. The multinominal 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 multinominal 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 – for all models in this thesis this will be the employment category of economic activity.

To understand multinominal logistic regression models further, suppose a dependent variable has M categories. One value is designated as the reference category. The probability membership in other categories is compared to the probability of membership in the reference category.

For a dependent variable with M categories, this requires the calculation of M-1 equations, one for each category relative to the reference category, to describe the relationship between the dependent variable and the independent variables.

Hence, for m=5,…,M,

Given that probabilities for the subsequent models are also being produced, for m=5,…,M,

And for the reference category,

Where relevant, 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.

The desire of this thesis to 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 ought to be selected 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 does not exist in a vacuum and 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 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 varinations of . The interpretation of a pseudo differs from its linera 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 numer of observations. The Cox-Snell can be calcualted for both linear or non-linear models – the equation is identical. As Allison states, this is more appropriately terms 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. 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, calculate the mean of the predicted probabilities of an event, then take the difference between the two means. 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 nominal logistic regression. Whilst statistically speaking the Tjur measure appears the most attractive, it is as of writing not applicable to the proposed multinominal logistic regression models planned in this thesis. 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, 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 multinominal 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.

## 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 cohort. The main statistical method used is a multinominal logistic regression. A sensitivity analysis for the cohort comparing a 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 discusses the relevant missing data literature before conducting a simulation to select the most appropriate handling missing data method. This is followed by a multiple imputation and comparison with complete records analysis. Part 1 ends with a conclusion of statistical and substantive findings. Part 2 and Part 3 duplicate this analysis excluding the handling missing data simulation. Each Part is contained and focused on the analysis of its own cohort. Part 4 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.

# The National Childhood Development Survey (1958): Youth Transitions in Education and Employment

"Let us be frank: most people have never had it so good". – Harold Macmillan (1957)

## Introduction

The transition from mandatory education into a form of economic activity is one of the first significant life choices an individual in the UK has to make. This choice forms a bridge between the phase of education and preparation for the world of work and the phase of active employment (Mayer, 2009). The notion of an individual’s ‘choice’ is a complicated affair prominent in youth transition literature (Micklewright, 1989; Schoon, 2010). The role of structural influences, such as social class, sex, and housing tenure, can potentially influence individuals' choices across their life course. These structural influences may provide opportunities or hinder individuals' decision-making when selecting 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. The cohort of individuals analysed in this chapter comes from the National Childhood Development Survey (NCDS). This cohort of 17,638 individuals was all born in the same week in March 1958 and grew up in an identical socio-historical context. The NCDS cohort left mandatory schooling at age 16 in 1974. During this period, the UK was experiencing a large-scale shift in its labour market. This shift consisted of a collapse of traditional manufacturing occupations with simultaneous growth in the service economy and a collapse in the youth labour market. The NCDS cohort experienced a society in which Harold Macmillan proclaimed so confidently that ‘’You never had it so good’’ (Hamnett et al., 1989). The 1960s saw the ‘white heat’ of Wilson’s technological revolution, transforming British society and its subsequent labour markets into a service-based, consumer-based economy (Hamnett et al., 1989). These large-scale economic impacts had reverberations linked to specific social stratification areas. Most apparent consisted of a relative decline in traditionally working-class occupations and a growth of women in the labour market.

The NCDS provides an ability to study the influences of structural inequalities within a socio-historical context upon an individual’s choice and opportunity post-mandatory schooling. The following chapter identifies five core transitionary pathways following post-mandatory education economic activity: employment, non-traditional education, school, training & apprenticeships, and unemployment & out of the labour force. These five transitionary pathways are analysed through structural inequalities influencing an individual’s decision-making in sorting into one of these five economic activities post-mandatory schooling. Figure 1.1 illustrates the transitionary pathways possible for the individuals within the NCDS cohort.

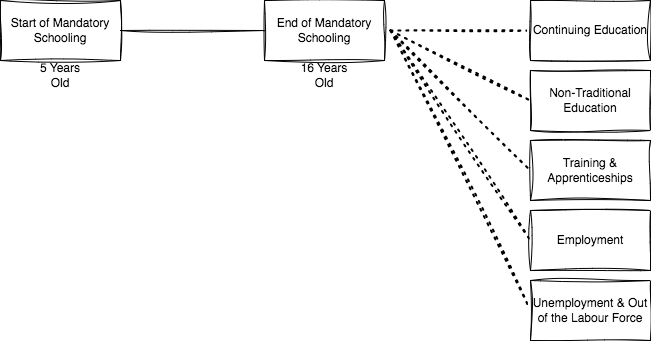


Figure 1.1 Transitional Pathways for NCDS Cohort

## Literature Review: NCDS Timeframe and Context

This section provides an overview of the literature on youth transitions. This literature review focuses on existing research outlining the school-to-work transition and examining the impact of structural inequalities on that transition. A significant focus is placed on the role of social class and sex. A broader focus also examines the nature of choice and opportunity for the NCDS cohort and how structural inequalities impact this. Major transition trajectories have been explored within this review. Trajectories into further education, apprenticeships, employment, and unemployment are identified. Each of these trajectories appears to have an element of structural inequalities influencing the outcomes of individuals. These empirical findings are expanded upon by reviewing the theoretical literature to provide a holistic overview of the school-to-work transition during the NCDS period.

Compulsory Schooling

During the NCDS cohort, 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).

### Story of transitions for NCDS youth

From the age of 16, NCDS youth had multiple pathways. Some would choose to enter the labour market and employment straight away. Others would seek to elongate their educational pathway by staying within school. Traditionally, this would mean joining a sixth-form college and taking Advanced levels (A’levels). During the NCDS timeframe, non-traditional[[3]](#footnote-3) educational pathways were also available. These were typically 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. Finally, some would enter a period of unemployment or opt to be out of the labour force.

Structural Influence on Choice and Opportunity for NCDS cohort

Individuals do not make choices in a vacuum. Given their biographical life course (Elder, 1994) up until the point of finishing mandatory education, stratifying influences play a role in what choices an individual is likely to select and what opportunities they can also participate in. Education is one such stratifying influence. Education regulates the individual by implementing age-graded barriers and hierarchical and time-related credentials. The type of education an individual experiences and the product of their educational attainment captures some of this structural inequality. As with education, social class is another potential stratifying influence. Class position also has a stratifying influence on society through occupations. The structure and hierarchy of occupations determine social position via segmentation and segregation, partly determined by previous employment and education systems (XXXX). Stratification is vital in the underlying choices and opportunities that influence individuals' societal pathways.

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 (Martin, Schoon and Ross, 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, heterogenous transitions would have a large number of clusters that were relatively evenly spread out. Results from Martin et al (2008) however convey the NCDS cohort has having 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) find that young people from less privileged backgrounds are more likely to be in training or apprenticeships. The declining state of apprenticeships and British manufacturing has a disproportional level of impact upon 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 individual had. 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 comparative 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 – the role of sex and social-class

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 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 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).

There was some slight decline in gender segregation within the labour force for the NCDS cohort 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 (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).

#### Educational Attainment and training

When looking at educational attainment (Holm and Jæger, 2011), 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 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).

## Data and Methods

Following from initial introduction of the NCDS data in Part 1, the following section will detail the NCDS cohort data and the methods used for subsequent analysis of that data. The analytical sample includes all cohort members that have some level of data on their economic activity in September when they are aged 16 years old. This sample will look at the impact of structural inequality in the form of educational attainment, sex, social class, and housing tenure and how that impacts their economic activity. By including all cohort members that have responses on the economic activity variable the total analytical sample for this analysis has an N=12,411.

Sample Size and Attrition

Table 1.2 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 2.1 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).

### 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. The NCDS cohort originated in 1958 (when participants were born) and continues today. Only data up until age 23 (wave 4) is considered for analysis.

From this variable selection variables related to economic activity, educational attainment, sex, social class, and housing tenure were selected for inclusion in subsequent analysis.

#### Economic Activity

The primary outcome variable of interest is the main economic activity of month 201 since birth – this translates to the month of September when all cohort members are aged 16. 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. This 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’[[4]](#footnote-4). 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[[5]](#footnote-5) |  |
| 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 recorded as follows: five of these collapsed into the unemployment & out of labour force category. One into a full-time education post-school category. One into a school category. Four into an employment category (using both Full-time and Part-time employment as well as FT+Other and PT+Other). One into missing data. The rest into a training/apprenticeship category – this was accomplished via a dominance approach, any combination of categories whereby training & apprenticeship were mentioned, they were given priority in coding over and above other categories – this means for example that those within the fulltime job + apprenticeship category were coded into the training & apprenticeship category over that of the employment category. This was justified by the belief that an apprenticeship program is a training advancement program but typically they did not pay the same as a full-time position, often times individuals would take on some form of employment alongside their training program – for this reason the training program is seen as the primary form of activity for individuals. The training/apprenticeship category contains apprenticeships, like the Training Opportunities Scheme (TOPs) training courses. The NCDS codes main economic activity in a way that creates five categories: employment, non-traditional education, school, training & apprenticeships, and unemployment & out of labour force. Main Economic Activity is determined based on whether that activity is conducted 21 hours or more per week for Education (Full and Part-time), a full-time job of more than 30 hours, a part-time job of less than 30 hours, unemployed if the respondent is actively searching for work, and out of the labour force if all else is not false.

Re-coding this variable was necessary to get at the nuance of some of the economic activity data. For example, much data was coded as full-time employment – including training schemes, apprenticeships, Technical and Vocational Educational Initiative (TVEI), and TOPs schemes.

Re-coding this variable translates into the five-category economic activity discussed earlier. These five categories breakdown into: employment, non-traditional education, school, training & apprenticeship, and unemployment & out of the labour force. Employment collapsed from part-time and full-time into a singular employment category due to the negligible sample size (n=39) of part-time work. Post-school education refers to credit received for completion of courses not in a school environment but given by an accredited college, trade school, workshops, etc. School is defined as anyone who, after completing mandatory schooling at age 16, decides to continue education at school for A-levels, etc. Training & Apprenticeship is defined by any individual undertaking a training, work, or apprenticeship related scheme. Finally, Unemployment and OLF are a combined category of all unemployed and those out of the labour force. Unfortunately, for sample size reasons, these two categories had to be combined for statistical power. However, it is recognised that there is a qualitative distinction between these two categories that may impact the statistical power presented within the models. A full breakdown can be found in table 1.8.

#### 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 – as seen in Table 1.4. Providing a clearly demarcated variable of less than five or five or more O’levels allows a clear differentiation between individuals that where academically successful at 16 and those that were less so.

Table 2.3 Educational Attainment Count Variable by Economic Activity

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Educational Attainment – Number of O’levels | | | | | | | | | | |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Nine or More | Total |
| Economic Activity |  |  |  |  |  |  |  |  |  |  |  |
| *Employment* | 3,024 | 586 | 368 | 251 | 234 | 181 | 113 | 70 | 35 | 25 | 4,887 |
| *Non-Traditional Education* | 161 | 116 | 110 | 100 | 98 | 145 | 132 | 79 | 46 | 66 | 1,053 |
| *School* | 159 | 93 | 100 | 135 | 203 | 368 | 399 | 562 | 612 | 1,081 | 3,712 |
| *Training/Apprenticeships* | 1,136 | 375 | 227 | 168 | 129 | 123 | 65 | 52 | 21 | 20 | 2,316 |
| *Unemployment and OLF* | 350 | 45 | 16 | 14 | 9 | 9 | 6 | 5 | 2 | 1 | 457 |
| Total | 4,830 | 1,215 | 821 | 668 | 673 | 826 | 715 | 768 | 716 | 1,193 | 12,425 |

#### Sex

Sex is a variable derived from sweep 0 [n622\_4]. Its inclusion for analysis is because, during the timeframe of the NCDS, 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.

#### 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).

#### 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 [n1152][[6]](#footnote-6).

#### Social Stratification: NS-SEC, CAMSIS, RGSC

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 – in particular 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 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 and the construction of socio-economic variables

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 (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 compared to SOC 2000 – by 10 years. This closeness in time should also present a more accurate portrayal of British society in comparison to the SOC 2000 codes. However, the NCDS model is not in isolation. 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 1.8.

Table 2.4 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 1.9.

Table 2.5 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 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[[7]](#footnote-7).

##### 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[[8]](#footnote-8) (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 aforementioned 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:

Table 2.6 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 |

The NCDS has been coded to codes taken in 1969 – these codes are in SOC2000 and SOC90 format (Gregg, 2012). Amongst the social stratification variables that are provided, 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 [N2SRGSC] 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.

##### 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” (Rose and Pevalin, 2001). 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 2.7 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 |

As with the RGSC, the occupational codes provided by Gregg (Gregg, 2012) allows the operationalisation of the complete NS-SEC class schema [N2SNSSEC]. This will provide the basis for comparison and sensitivity analysis of socio-economic measures within this chapter. The following analytical variables within the NS-SEC have been broken down with example occupations to aid in interpretation within subsequent models in Table 1.7.

Table 2.8 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) argues that individuals are embedded within socially moderated spaces and networks within which they engage in various social and economic interactions, different from interactions with persons more distant from these networks (Stewart, Prandy and Blackburn, 1973, 1980). In other words, 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 2.9 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 |

CAMSIS was thus coded using SOC codes [N2SSOC90 and N2SSOC00] from Greggs occupational coding dataset (2012). Like NS-SEC, details on the employment status of individuals’ fathers were unavailable, so a ‘simplified CAMSIS’ was constructed. After this recoding, a comparison was made between this recoding and the original CAMSIS variable constructed by Gregg by multiplying the original CAMSIS values by 10. The former was much closer to the mean of 50, s.t.d of 15, which is expected from CAMSIS, as seen in Table 1.8. Due to the CAMSIS scale used in this analysis being the simplified scale, it is not surprising that it does not match exactly to the expected mean of 50, s.t.d of 15 threshold that the full scale does.

## Descriptive Statistics

Table 1.10 shows the frequencies and summary statistics for the NCDS. Overall, 38.25 per cent of the sample is in full-time employment. Whilst 30.33 per cent remain in school, 8.85 per cent moved on to full-time post-school education. Unemployment and being out of the labour force make up 3.07 per cent. Finally, 19.51 per cent of the sample are in some training or apprenticeship scheme.

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 a relatively 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 very 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 comparative to their SOC 2000 counterparts and NS-SEC 3, 5, 6, and 7 increased comparative to their SOC 2000 counterparts.

Looking again at table 1.10, 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 2.10 Descriptive Statistics for Economic Activity Model

|  |  |  |
| --- | --- | --- |
|  | n | % |
| Economic Activity of Respondent on September when they are 16 |  |  |
| *Employment* | 3,217 | 38.25% |
| Non-Traditional Education | 744 | 8.85% |
| *School* | 2,551 | 30.33% |
| *Training/Apprenticeships* | 1,641 | 19.51% |
| *Unemployment and OLF* | 258 | 3.07% |
| Educational Attainment O-levels |  |  |
| *Less than 5 O-Levels* | 5,426 | 64.51% |
| *Five or more 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 2.11 Descriptive Statistics by Economic Activity

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Economic Activity of Respondent on September when they are 16 n(%) | | | | | |
|  | Employment | Non-Traditional Education | School | Training/Apprenticeships | Unemployment and OLF | Total |
| Educational Attainment O-levels |  |  |  |  |  |  |
| *Less than 5 O-Levels* | 2922 (90.83%) | 403 (54.17%) | 435 (17.05%) | 1424 (86.78%) | 242 (93.80%) | 5426 (64.51%) |
| *Five or more 5 O-Levels* | 295 (9.17%) | 341 (45.83%) | 2116 (82.95%) | 217 (13.22%) | 16 (6.20%) | 2985 (35.49%) |
| Sex of Respondent |  |  |  |  |  |  |
| *Female* | 1955 (60.77%) | 521 (70.03%) | 1281 (50.22%) | 303 (18.46%) | 155 (60.08%) | 4215 (50.11%) |
| *Male* | 1262 (39.23%) | 223 (29.97%) | 1270 (49.78%) | 1338 (81.54%) | 103 (39.92%) | 4196 (49.89%) |
| Housing Tenure of Respondent when Child |  |  |  |  |  |  |
| *Own Home* | 1073 (33.35%) | 461 (61.96%) | 1734 (67.97%) | 705 (42.96%) | 72 (27.91%) | 4045 (48.09%) |
| *Don't Own Home* | 2144 (66.65%) | 283 (38.04%) | 817 (32.03%) | 936 (57.04%) | 186 (72.09%) | 4366 (51.91%) |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |  |  |  |  |
| *Large Employers and higher managerial occupations* | 49 (1.52%) | 33 (4.44%) | 146 (5.72%) | 30 (1.83%) | 3 (1.16%) | 261 (3.10%) |
| *Higher professional occupations* | 43 (1.34%) | 37 (4.97%) | 291 (11.41%) | 38 (2.32%) | 1 (0.39%) | 410 (4.87%) |
| *Lower Managerial and professional occupations* | 213 (6.62%) | 131 (17.61%) | 544 (21.32%) | 129 (7.86%) | 21 (8.14%) | 1038 (12.34%) |
| *Intermediate occupations* | 227 (7.06%) | 93 (12.50%) | 354 (13.88%) | 119 (7.25%) | 12 (4.65%) | 805 (9.57%) |
| *Small employers and own account workers* | 401 (12.47%) | 84 (11.29%) | 269 (10.54%) | 241 (14.69%) | 29 (11.24%) | 1024 (12.17%) |
| *Lower supervisory and technical occupations* | 546 (16.97%) | 12 (16.94%) | 354 (13.88%) | 314 (19.13%) | 32 (12.40%) | 1372 (16.31%) |
| *Semi-routine occupations* | 680 (21.14%) | 97 (13.04%) | 305 (11.96%) | 340 (20.72%) | 63 (24.42%) | 1485 (17.66%) |
| *Routine occupations* | 1058 (32.89%) | 143 (19.22%) | 288 (11.29%) | 430 (26.20%) | 97 (37.60%) | 2016 (23.97%) |
| RGSC Social Class of Father when Respondent Child SOC2000 |  |  |  |  |  |  |
| *Professional* | 35 (1.09%) | 33 (4.44%) | 257 (10.07%) | 36 (2.19%) | 1 (0.39%) | 362 (4.30%) |
| *Managerial and Technical* | 415 (12.90%) | 199 (26.75%) | 836 (32.77%) | 235 (14.32%) | 35 (13.57%) | 1720 (20.45%) |
| *Skilled non-manual* | 253 (7.86%) | 111 (14.92%) | 378 (14.82%) | 154 (9.38%) | 9 (3.49%) | 905 (10.76%) |
| *Skilled manual* | 1538 (47.81%) | 285 (38.31%) | 759 (29.75%) | 806 (49.12%) | 113 (43.80%) | 3501 (41.62%) |
| *Partly skilled* | 567 (17.63%) | 81 (10.89%) | 226 (8.86%) | 272 (16.58%) | 59 (22.87%) | 1205 (14.33%) |
| *Unskilled* | 409 (12.71%) | 35 (4.70%) | 95 (3.72%) | 138 (8.41%) | 41 (15.89%) | 718 (8.54%) |
| NS-SEC Social Class of Father when Respondent Child SOC90 |  |  |  |  |  |  |
| *Large Employers and higher managerial occupations* | 2 (0.06%) | 1 (0.13%) | 5 (0.20%) | 1 (0.06%) | 0 (0.00%) | 9 (0.11%) |
| *Higher professional occupations* | 45 (1.40%) | 31 (4.17%) | 236 (9.25%) | 32 (1.95%) | 2 (0.78%) | 346 (4.11%) |
| *Lower Managerial and professional occupations* | 122 (3.79%) | 92 (12.37%) | 376 (14.74%) | 86 (5.24%) | 13 (5.04%) | 689 (8.19%) |
| *Intermediate occupations* | 202 (6.28%) | 110 (14.78%) | 428 (16.78%) | 122 (7.43%) | 8 (3.10%) | 870 (10.34%) |
| *Small employers and own account workers* | 282 (8.77%) | 54 (7.26%) | 186 (7.29%) | 136 (8.29%) | 20 (7.75%) | 678 (8.06%) |
| *Lower supervisory and technical occupations* | 529 (16.44%) | 138 (18.55%) | 386 (15.13%) | 329 (20.05%) | 32 (12.40%) | 1414 (16.81%) |
| *Semi-routine occupations* | 855 (26.58%) | 154 (20.70%) | 551 (21.60%) | 430 (26.20%) | 70 (27.13%) | 2060 (24.49%) |
| *Routine occupations* | 1180 (36.68%) | 164 (22.04%) | 383 (15.01%) | 505 (30.77%) | 113 (43.80%) | 2345 (27.88%) |
| RGSC Social Class of Father when Respondent Child SOC90 |  |  |  |  |  |  |
| *Professional* | 36 (1.12%) | 29 (3.90%) | 208 (8.15%) | 28 (1.71%) | 3 (1.16%) | 304 (3.61%) |
| *Managerial and Technical* | 106 (3.29%) | 79 (10.62%) | 381 (14.94%) | 73 (4.45%) | 12 (4.65%) | 651 (7.74%) |
| *Skilled non-manual* | 295 (9.17%) | 149 (20.03%) | 504 (19.76%) | 173 (10.54%) | 8 (3.10%) | 1129 (13.42%) |
| *Skilled manual* | 1235 (38.39%) | 172 (23.12%) | 519 (20.34%) | 592 (36.08%) | 83 (32.17%) | 2601 (30.92%) |
| *Partly skilled* | 1159 (36.03%) | 267 (35.89%) | 781 (30.62%) | 620 (37.78%) | 113 (43.80%) | 2940 (34.95%) |
| *Unskilled* | 386 (12.00%) | 48 (6.45%) | 158 (6.19%) | 155 (9.45%) | 39 (15.12%) | 786 (9.34%) |
| CAMSIS Score of Father when Respondent Child SOC2000 | 40.09 (11.12) | 47.55 (13.37) | 51.88 (14.71) | 41.62 (11.54) | 38.26 (10.66) | 44.57 (13.63) |
| CAMSIS Score of Father when Respondent Child SOC90 | 38.75 (10.34) | 43.94 (13.35) | 47.73 (14.71) | 39.60 (10.94) | 36.84 (10.00) | 42.04 (12.84) |
| N | 3217 (38.25%) | 744 (8.85%) | 2551 (30.33%) | 1641 (19.51%) | 258 (3.07%) | 8411 (100.00%) |
| Data Source: NCDS [Sweeps 0-4] | | | | | | |

Descriptive statistics by economic activity

From Table 1.11, some observations can be made. An individual’s educational attainment is widely different when stratified by their economic activity. Those that enter employment have a split of 90.83 per cent having achieved less than five O’levels compared to 9.17 per cent of their peers that achieved five or more O'levels. The reverse is true for those who stayed at school, whereby 82.95 per cent of individuals achieved five or more levels. The split of educational attainment is almost identical for those who entered employment and those who entered a training and apprenticeship scheme, as well as those who entered a period of unemployment or out of the labour force. It is only non-traditional education that has a somewhat even split between individuals who have achieved five or more levels and those who have not.

From observing the descriptive statistics, economic activity is stratified heavily by sex. Whilst there is an even split of women and men staying on in school, those who decide to enter training and apprenticeship schemes are dominated by men (81.54 per cent). Comparatively, those who enter non-traditional education are primarily women (70.03 per cent). Women are also a majority in entering employment and being unemployed or out of the labour force. Men are only the majority in one economic activity category – training and apprenticeships.

Those who lived with parents who did not own their own homes make up the majority (72.09 per cent) of individuals who are unemployed and out of the labour force category. Most individuals who chose to stay within school or go to non-traditional education also had parents who owned their own home (at rates of 67.97 per cent and 61.96 per cent, respectively). For those who chose employment or training and apprenticeship schemes, most came from parents who did not own their own home (at 66.65 per cent and 57.04 per cent, respectively.

Looking at NS-SEC for SOC 2000 construction, the largest concentration of NS-SEC 1.1, 1.2, 2 and 3 is concentrated within the school category at 5.72 per cent, 11.41 per cent, 21.32 per cent, and 13.88 per cent respectively. The lowest concentrations of 1.1, 1.2, and 3 are concentrated within the unemployment and out of the labour force category at 1.16 and per cent, 0.39 per cent, and 4.65 per cent respectively. Unlike NS-SEC 1.1, 1.2, and 3, NS-SEC 2 shares its lowest concentration is within employment at 6.62 per cent. The largest concentration of NS-SEC 4 and 5 are within training and apprenticeship programs at 14.69 per cent and 19.13 per cent, respectively; however, they deviate from each other concerning their lowest concentration. For NS-SEC 4, the lowest concentration is within schools at 10.54 per cent, whereas for NS-SEC 5, it is within unemployment and out of the labour force at 12.40 per cent. NS-SEC 6 and 7 share the highest concentration of individuals within unemployment and out of the labour force at 24.42 per cent and 37.60 per cent, respectively, and a shared lowest concentration within school at 11.96 per cent and 11.29 per cent, respectively. Looking at NS-SEC within each economic activity, there is a linear increase in individuals participating in employment as NS-SEC increases (from 1.1-7). When looking at non-traditional education and comparing the per cent of each NS-SEC category to their total, it is evident that NS-SEC 1.1, 2, and 3 are overrepresented in this economic activity, whilst NS-SEC 4-7 are underrepresented. For schools, NS-SEC 1.1-3 are overrepresented in this economic activity outcome, whereas NS-SEC 4-7 are underrepresented in this category. The exact reverse is true regarding the training and apprenticeship category. The only NS-SEC categories overrepresented in the unemployment and out of the labour force category are NS-SEC 6 and 7, with all else being underrepresented.

Compared to the SOC 90 construction of NS-SEC, NS-SEC 1.1, 1.2, 2, and 3 share the largest and smallest concentrations with their SOC 2000 counterpart. NS-SEC 4 deviates however, with its largest concentration being in employment at 8.77 per cent compared to SOC 2000 NS-SEC being in training & apprenticeships at 14.69 per cent. The smallest concentration also deviates, with SOC 90 NS-SEC 4 being in non-traditional education at 7.26 per cent and the SOC 2000 alternative being in school at 10.54 per cent. Both SOC 2000 and SOC 90 constructions of NS-SEC concur on the largest and smallest concentrations within NS-SEC 5. Whilst both concur that the largest concentration of individuals in NS-SEC 6 are situated within unemployment & OLF, they once again diverge in relation to the smallest concentration. The SOC 2000 construction of NS-SEC 6 states the smallest concentration is within school at 11.96 per cent but the SOC 90 construction states it is in fact within non-traditional education at 20.70 per cent. Both concur that NS-SEC 7 has the largest concentration in unemployment & OLF and the smallest concentration in school.

For the RGSC schema for SOC 2000 construction, an explicit manual/non-manual divide becomes apparent when looking at Table 1.9. Delineating RGSC 1-3 as non-manual and RGSC 4-6 as manual, for employment, training and apprenticeships, and unemployment and out of the labour force economic activities, those in manual occupations are overrepresented, and those in non-manual are underrepresented. The reverse for non-traditional education and school is genuine, whereby manual occupations are underrepresented, and non-manual occupations are overrepresented. The same general trends are present for the SOC 90 construction of RGSC.

For CAMSIS for SOC 2000 construction, with a base total mean of 42.12, like RGSC, there is a delineation between categories of economic activity relating to employment, training and apprenticeships, and unemployment and out of the labour force on the one hand, and on the other, non-traditional education and school. Those in the former group have a CAMSIS mean below the total, and those in the latter have a CAMSIS mean above the total. This trend is replicated using the SOC 90 construction of CAMSIS.

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. Starting with table 1.12 whilst 100 per cent of NS-SEC 1.1 for SOC 90 is also within the SOC 2000 construction there is a large amount of individuals that are coded from 1.2-6 in the SOC 90 construction that are otherwise coded as 1.1 in the SOC 2000 construction. Across all NS-SEC categories the majority of individuals are sorted into the same NS-SEC category in the SOC 2000 as well as SOC 90 construction. The lowest share of same category sorting amounts to 60.69 per cent for NS-SEC 3. The highest share of the same category sorting amounts to NS-SEC 1.1 at 100 per cent.

Moving on to table 1.13, like NS-SEC there are a number of deviations from the SOC 90 coding compared to the SOC 2000 construction. If looking at RGSC with a manual/non-manual divide between 3NM and 3M, a majority of individuals are always sorted into the appropriate manual/non-manual divide.

Table 1.14 demonstrates that CAMSIS under the SOC 2000 construction has both a higher mean and a higher standard deviation compared to the SOC 90 counterpart. However, whilst SOC 2000 is higher, the difference is relatively marginal, suggesting that the substantive results in any model will be close to one another.

Table 2.12 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 |
| 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%) |
| 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%) |
| Data Source: NCDS [Sweeps 0-4] | | | | | | | | | |

Table 2.13 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 |
| 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%) |
| N | 304 (3.61%) | 651 (7.74%) | 1129 (13.42%) | 2601 (30.92%) | 2940 (34.95%) | 786 (9.34%) | 8411 (100.00%) |
| Data Source: NCDS [Sweeps 0-4] | | | | | | | |

Table 2.14 Descriptive Statistics comparing CAMSIS by SOC2000 and SOC90 codes

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

## Modelling Main Economic Activity

The primary outcome variable is the main economic activity of individuals in September of 1974. This is the first-month individuals were in when they received their O’level results after mandatory schooling. The first set of analyses estimates a multinomial logistic regression model with NS-SEC as the chosen social stratification measure (RGSC and CAMSIS will be introduced in a sensitivity analysis later). Table 1.15 details the deviance, change in deviance, change in degrees of freedom, and McFadden’s Adjusted Pseudo , AIC, and BIC measures to compare the null model with models of one explanatory variable. Table 1.10 details the exact statistics but through a sequential building of the null model with each subsequent independent variable added.

This model has been tested 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.

The model output uses the reference category of the school. The schooling category contrasts with all other economic activity categories because it has the most significant barrier to entry; continuing schooling expects previous educational merit. School as a reference category is sociologically compelling. Contrasting school with other economic activity destinations like employment or apprenticeships 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[[9]](#footnote-9).

Table 2.15 Goodness-of-fit summaries for explanatory variables and Economic Activity

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 23040.88 | - | - | - | 23048.88 | 23077.03 |
| Null Model + Educational Attainment | 18829.53 | 4211.35 | 4 | 0.18 | 18845.83 | 18901.83 |
| Null Model + Sex | 22052.16 | 988.72 | 4 | 0.04 | 22068.16 | 22124.46 |
| Null Model + Tenure | 22224.78 | 816.11 | 4 | 0.04 | 22240.77 | 22297.07 |
| Null Model + NS-SEC (SOC 2000) | 21772.23 | 1268.65 | 28 | 0.06 | 21836.23 | 22061.42 |

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

Table 2.16 Model building goodness-of-fit summaries for multinominal logistic regression model of Economic Activity

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 23040.88 | - | - | - | 23048.88 | 23077.03 |
| Null Model + Educational Attainment | 18829.53 | 4211.35 | 4 | 0.18 | 18845.83 | 18901.83 |
| Null Model + Educational Attainment + Sex | 17843.77 | 985.76 | 4 | 0.23 | 17867.77 | 17952.22 |
| Null Model + Educational Attainment + Sex + Tenure | 17602.92 | 240.85 | 4 | 0.24 | 17634.92 | 17747.51 |
| Null Model + Educational Attainment + Sex + Tenure + NS-SEC (SOC 2000) | 17343.50 | 259.42 | 28 | 0.25 | 17431.50 | 17741.14 |

The model fit statistics demonstrate that there are typically distributed residuals and that the model is correctly specified. Table 1.16 suggests that deviance is reduced by 5,697.38 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 0.25. In other words, the full model explains 25 per cent of the variance of economic activity, leaving 75 per cent unexplained. The following analysis with the full model is a complete records analysis with 8,411 observations.

Before discussing this model's results, a discussion on interpretation must be had. When dealing with multinomial 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. However, odds ratios have their issues, which result in an inability to compare across models and datasets, even if they have the exact model specification (ibid). Sometimes, odds ratios cannot be compared and interpreted within a model (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. Due to these issues, both log odds and odds ratios provide an underwhelming desire to use them to interpret multinominal logistic models beyond establishing primary substantive effects of ‘higher’ and ‘lower’ (Gayle and Lambert, 2009). The popular 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 being in an economic activity category over school, holding all other variables at their observed values. The rationale for interpreting multinomial logistic models using average marginal effects is based on the fact that the marginal effect is less sensitive to changes in model specification than the odds ratio, and the average marginal effect can be either positive or negative. Finally, average marginal effects for subgroups (like social class) can differ, leading to different implications and interpretations (Norton and Dowd, 2018).

For subsequent analysis, log odds will be presented for primary substantive effects, with average marginal effects used alongside as a complement to interpretation.

The results of the multinomial logistic regression model are reported in Table 1.17. 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. 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 1.17. 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’. 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[[10]](#footnote-10) (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. Recoding the outcome variable three times so that each outcome category of the outcome variable is coded once as '1' means that the QV statistics can be procured for each category and then combined later. QV statistics and QV graphs for each category of the outcome variable can be produced within the Stata environment. QV statistics are reported alongside log odds and average marginal effects, whilst QV graphs are reported below.

The output for employment demonstrates that individuals who received five or more O’levels have decreased log odds of employment over the school. Using average marginal effects, there is a 39 per cent decreased probability for an individual to be employed over school if they received five or more O’levels. Educational attainment has the most substantial impact on an individual’s choice to be employed over school. Men had a decreased log odds of being in employment over school. Regarding average marginal effects, this translates to a 17 per cent decreased probability that men would be employed over school. Individuals with parents who did not own their own home when a child had increased log odds of employment compared to school. Translated into average marginal effects, this represents an 8 per cent increased probability for an individual to be employed over school if they lived in a household where their parents did not own their own home as a child. Using NS-SEC 2 as a reference category, every other NS-SEC category except NS-SEC 1.2 has an increased log odds of employment over school compared to NS-SEC 2 – NS-SEC 1.2 has a decreased log odds of being in employment over school compared to the reference category.

Translated into average marginal effects individuals within NS-SEC 1.2 social origins have a 5 per cent decreased probability of being in employment over school compared to individuals with NS-SEC 2 social origins. For individuals in NS-SEC 3 social origins onwards there is a general trend of increasing probability for individuals in lower NS-SEC social origins to be in employment over school compared to their NS-SEC 2 peers. The largest of these effects refers to individuals residing in NS-SEC 7 social origins, whereby there is a 13 per cent increased probability for individuals whose social origins are NS-SEC 7 to be in employment over school when compared to the reference category. For a full breakdown of the marginal effects of NS-SEC in the employment category, see Figure 1.2 and its explanation.

The output for non-traditional education demonstrates that individuals who received five or more O’levels had increased log odds of being in non-traditional education over school. Using average marginal effects, there is a 3 per cent increased probability for an individual to be in non-traditional education over employment if they received five or more O’levels. Men had a decreased log odds of being in non-traditional education over school than women, or in terms of average marginal effects, a 7 per cent decreased probability of being in non-traditional education over employment if the individual is a man. Sex is the single most substantial impact on an individual’s choice to enter non-traditional education. Housing tenure was found to be not statistically significant and thus will not be interpretated. Moving on to NS-SEC, individuals that have a social origins of NS-SEC 1.2 had a decreased log odds of being in non-traditional education over school compared to their NS-SEC 2 social origins peers. This translates to a 1 per cent decreased probability in terms of average marginal effects. Likewise, those individuals that occupy a NS-SEC 7 social origins had a decreased log odds of being in non-traditional education over school compared to their NS-SEC 2 peers. Translated to average marginal effects this results in a 3 per cent decreased probability.

The output for training & apprenticeships demonstrates that individuals that received five or more O’levels had a decreased log odds of being in training & apprenticeships over school, in terms of average marginal effects, this corresponds to a decreased probability of 17 per cent. Men compared with women had an increased log odds of being in training & apprenticeships over employment, or a 24 per cent increased probability. Sex is the single strongest predictor of whether an individual chooses to enter training and apprenticeship schemes over school. Results suggest that individuals that did not own their own home compared to those that did have a decreased log odds of being in training & apprenticeships over school. As this corresponds to average marginal effects, there is a 1 per cent decrease in probability of being in training & apprenticeships over employment for an individual that lives in a home that their parents do not own over people that do. Results also suggest that individuals whose social origins were NS-SEC 4-7 had an increased log odds of being in training & apprenticeships over school compared to their NS-SEC 2 social origin peers. The largest per cent probability of being in training & apprenticeships over school resides within NS-SEC 4 with a 5 per cent increased probability in comparison with NS-SEC 2. This is of little surprise considering NS-SEC 4 consists of small employers and own account workers – in other words, the self-employed. It makes sense that children of the self-employed would have a higher probability to enter training & apprenticeship programs which may lead to self-employed in the future.

The output for unemployment & OLF demonstrates that individuals who received five or more O’levels had decreased log odds of being unemployed and out of the labour force compared to those in school. This translates into a 3 per cent decreased probability of being unemployed and out of the labour force compared to being in school. Men are less likely to be unemployed or out of the labour force than women, with decreased log odds translating to a 1 per cent decreased probability. Individuals from households where their parents do not own their own home have increased odds of being unemployed and out of the labour force compared to being in school. Regarding average marginal effects, this corresponds to a 1 per cent increased probability of being unemployed and out of the labour force compared to being in school. Moving on to NS-SEC, whilst NS-SEC 1.2 has a decreased log odds of being in unemployment & OLF over school compared to NS-SEC 2 – translated to a 3 per cent decreased probability – NS-SEC 6 has an increased log odds, though when translated to average marginal effects results in a 0 per cent increased probability.

Table 2.17 Mlogit of Economic Activity

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NS-SEC | | | Average Marginal Effects | | Quasi-variance | | |
| Economic Activity: ‘School’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **S.E.** | **LCI** | **UCI** |
| Employment |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Five or More O’levels* | -3.58 | (0.08) | \*\*\* | -0.39 | (0.01) | (.) | (.) | (.) |
| Sex |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Male* | -0.40 | (0.07) | \*\*\* | -0.17 | (0.01) | (.) | (.) | (.) |
| Housing Tenure |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Do not Own Home* | 0.68 | (0.08) | \*\*\* | 0.08 | (0.01) | (.) | (.) | (.) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |
| *1.1* | 0.08 | (0.23) |  | 0.02 | (0.03) | 0.21 | -0.34 | 0.50 |
| *1.2* | -0.67 | (0.22) | \*\* | -0.05 | (0.03) | 0.20 | -1.06 | -0.27 |
| *2* | Ref. | (.) |  | (.) | (.) | 0.10 | -0.21 | 0.21 |
| *3* | 0.36 | (0.15) | \* | 0.05 | (0.02) | 0.11 | 0.14 | 0.58 |
| *4* | 0.96 | (0.14) | \*\*\* | 0.08 | (0.02) | 0.10 | 0.75 | 1.16 |
| *5* | 0.87 | (0.14) | \*\*\* | 0.07 | (0.02) | 0.09 | 0.69 | 1.04 |
| *6* | 0.91 | (0.14) | \*\*\* | 0.09 | (0.02) | 0.09 | 0.73 | 1.09 |
| *7* | 1.36 | (0.13) | \*\*\* | 0.13 | (0.02) | 0.08 | 1.19 | 1.53 |
| Intercept | 0.89 | (0.12) | \*\*\* | (.) | (.) | (.) | (.) | (.) |
|  |  |  |  |  |  |  |  |  |
| Non-Traditional Education |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Five or More O’levels* | -1.71 | (0.09) | \*\*\* | 0.03 | (0.01) | (.) | (.) | (.) |
| Sex |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Male* | -0.84 | (0.09) | \*\*\* | -0.07 | (0.01) | (.) | (.) | (.) |
| Housing Tenure |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Do not Own Home* | -0.11 | (0.10) |  | -0.04 | (0.01) | (.) | (.) | (.) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |
| *1.1* | 0.02 | (0.23) |  | 0.00 | (0.02) | 0.20 | -0.39 | 0.43 |
| *1.2* | -0.52 | (0.21) | \* | -0.01 | (0.02) | 0.18 | -0.89 | -0.15 |
| *2* | Ref. | (.) |  | (.) | (.) | 0.10 | -0.21 | 0.21 |
| *3* | 0.08 | (0.16) |  | -0.01 | (0.01) | 0.12 | -0.17 | 0.32 |
| *4* | 0.15 | (0.17) |  | -0.03 | (0.01) | 0.13 | -0.11 | 0.41 |
| *5* | 0.24 | (0.15) |  | -0.02 | (0.01) | 0.11 | 0.02 | 0.46 |
| *6* | -0.02 | (0.16) |  | -0.04 | (0.01) | 0.12 | -0.27 | 0.23 |
| *7* | 0.44 | (0.15) | \*\* | -0.03 | (0.01) | 0.11 | 0.22 | 0.66 |
| Intercept | 0.18 | (0.13) |  | (.) | (.) | (.) | (.) | (.) |
|  |  |  |  |  |  |  |  |  |
| School | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
|  |  |  |  |  |  |  |  |  |
| Training & Apprenticeships |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Five or More O’levels* | -3.24 | (0.09) | \*\*\* | -0.17 | (0.01) | (.) | (.) | (.) |
| Sex |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Male* | 1.52 | (0.09) | \*\*\* | 0.24 | (0.01) | (.) | (.) | (.) |
| Housing Tenure |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Do not Own Home* | 0.38 | (0.08) | \*\*\* | -0.01 | (0.01) | (.) | (.) | (.) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |
| *1.1* | -0.12 | (0.26) |  | -0.02 | (0.03) | 0.24 | -0.60 | 0.36 |
| *1.2* | -0.45 | (0.23) |  | 0.00 | (0.03) | 0.20 | -0.86 | -0.04 |
| *2* | Ref. | (.) |  | (.) | (.) | 0.12 | -0.24 | 0.24 |
| *3* | 0.19 | (0.17) |  | -0.00 | (0.02) | 0.13 | -0.07 | 0.45 |
| *4* | 0.97 | (0.16) | \*\*\* | 0.05 | (0.02) | 0.11 | 0.76 | 1.20 |
| *5* | 0.90 | (0.15) | \*\*\* | 0.05 | (0.02) | 0.10 | 0.71 | 1.09 |
| *6* | 0.84 | (0.15) | \*\*\* | 0.04 | (0.02) | 0.10 | 0.65 | 1.04 |
| *7* | 1.08 | (0.15) | \*\*\* | 0.02 | (0.02) | 0.09 | 0.90 | 1.27 |
| Intercept | -0.72 | (0.14) | \*\*\* | (.) | (.) | (.) | (.) | (.) |
|  |  |  |  |  |  |  |  |  |
| Unemployment & Out of Labour Force |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Five or More O’levels* | -3.94 | (0.27) | \*\*\* | -0.03 | (0.00) | (.) | (.) | (.) |
| Sex |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Male* | -0.37 | (0.14) | \*\* | -0.01 | (0.00) | (.) | (.) | (.) |
| Housing Tenure |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Do not Own Home* | 0.8 | (0.16) | \*\*\* | 0.01 | (0.00) | (.) | (.) | (.) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |
| *1.1* | -0.35 | (0.67) |  | -0.01 | (0.02) | 0.60 | -1.57 | 0.86 |
| *1.2* | -2.08 | (1.03) | \* | -0.03 | (0.01) | 1.01 | -4.12 | -0.04 |
| *2* | Ref. | (.) |  | (.) | (.) | 0.24 | -0.48 | 0.48 |
| *3* | -0.29 | (0.38) |  | -0.01 | (0.01) | 0.30 | -0.90 | 0.33 |
| *4* | 0.60 | (0.31) |  | -0.00 | (0.01) | 0.21 | 0.19 | 1.02 |
| *5* | 0.29 | (0.30) |  | -0.01 | (0.01) | 0.19 | -0.10 | 0.68 |
| *6* | 0.77 | (0.28) | \*\* | 0.00 | (0.01) | 0.15 | 0.47 | 1.07 |
| *7* | 1.20 | (0.27) | \*\*\* | 0.01 | (0.01) | 0.13 | 0.94 | 1.46 |
| Intercept | -1.48 | (0.25) | \*\*\* | (.) | (.) | (.) | (.) | (.) |
|  |  |  |  |  |  |  |  |  |
| Number of observations | 8411 | | | | | | | |
| McFadden’s | 0.25 | | | | | | | |
| McFadden’s Adjusted Pseudo | 0.24 | | | | | | | |
| Cox-Snell Pseudo | 0.49 | | | | | | | |
| Nagelkerke Pseudo | 0.53 | | | | | | | |
| AIC | 17431.50 | | | | | | | |
| BIC | 17741.14 | | | | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: Complete Records Analysis | | | | | | | | |

To understand this in a more manageable format, each variable is graphically visualised with their predicted probabilities. This allows for a more intuitive understanding of the multinominal logistic regression and provides a different outlook for interpretation. 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.

Focusing on NS-SEC, figure 1.2 depicts the predicted probabilities at means of economic activity. Predicted probabilities for each economic activity category are presented. With the exception of 1.1-1.2, where there is a slight decline in people going into employment, there is a general upward trend for individuals to be employed from NS-SEC 1.2-7. For non-traditional education, there is a relative flatline across the NS-SEC schema. For the school category, there is an almost identical reverse picture of what happens with individuals sorting into the employment category. There is an increase from 1.1-1.2, and from that point onwards, there is a decline from NS-SEC 1.2-7. Training & apprenticeships has a unique dichotomous state. NS-SEC 1.1-3 are split into one group, and NS-SEC 4-7 into another. The former group has a decreased likelihood of entering this category compared to the latter. The peak of this group comes from NS-SEC 4, characterised by small employers and own account workers. The results from this model reaffirm prior research on the division of opportunity within skilled and non-skilled labour as it pertains to training & apprenticeships schemes. Skilled workers in NS-SEC 4 have the highest probability of being in a training and apprenticeship scheme compared to all other categories in the NS-SEC. This is not entirely surprising, given the socio-historical context of the NCDS. Apprenticeship & training schemes were heavily influenced by skilled worker occupations (Booth and Satchell, 1994). Unemployment has a relative flatline effect.

Figure 1.2 Predicted Probabilities of Economic Activity by NS-SEC

A graph showing the number of people in the united states

Description automatically generated

Moving on to Sex, figure 1.3 demonstrates that women are more likely than men to enter employment post-mandatory schooling. This is also true for less traditional non-traditional education and schools (though comparatively, the effect sizes are much smaller for schools than for non-traditional education). There is a relatively identical number of men compared to women entering school post-mandatory education. The most considerable sex-based effect relates to men being more likely to enter training & apprenticeship pathways compared to women. Given that the increase in soft skilled labour had only recently begun (Guinea-Martin and Elliott, 2008) under the NCDS cohort, alongside the predisposition for apprenticeship programs to be focused primarily upon skilled manual labour (Booth and Satchell, 1994), it is somewhat understandable as to why this effect size is as large as it is.

Figure 1.3 Predicted Probabilities of Economic Activity by Sex

A graph showing the number of people in the world

Description automatically generated with medium confidence

Moving on to educational attainment, figure 1.4 demonstrates that effect sizes are larger when discussing attainment over other variables. Those who gained less than five O’levels at school were more likely to enter employment than their peers who did gain five or more O’levels. Predictably, those individuals who gained five or more O’levels were more likely to enter school pathways and less likely to enter employment pathways compared to their peers who received less than five O’levels. Those who earned less than five O’levels were more likely than their academically successful peers to enter training and apprenticeship and non-traditional education pathways. This suggests that for the NCDS cohort, even those who did not do well academically, there were various pathways above and beyond entering employment straight after mandatory schooling ended. For those who did want to continue their education in some manner, pathways were available to them that were not rigidly restricted based on academic merit at earlier life stages.

Figure 1.4 Predicted Probabilities of Economic Activity by Educational Attainment

A graph showing the number of probabilities of economic activity

Description automatically generated

Moving on to housing tenure, figure 1.5 demonstrates that whilst substantive findings exist, the effect sizes compared to other variables are the smallest. Those who grew up in households that owned their own home were less likely to enter employment straight after mandatory schooling ended and more likely to enter some form of education – either non-traditional education or school-based pathways.

Figure 1.5 Predicted Probabilities of Economic Activity by Housing Tenure

A graph showing the number of individuals in the economic activity

Description automatically generated with medium confidence

Alongside the graphical presentation of predicted probabilities, the following figures also visualize the log odds of NS-SEC within each outcome category (except the reference category of school) alongside quasi-variance statistics to overcome the reference category problem. The underlying trend amongst all quasi-variance figures 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.

A graph showing the number of numbers and the number of bars

Description automatically generated with medium confidence

Figure 1.6 Log Odds versus Quasi-variance statistics of individuals being in School over Employment

A graph with red and black lines

Description automatically generated

Figure 1.7 Log Odds versus Quasi-variance statistics of individuals being in School over Unemployment & OLF

A graph with red and black lines

Description automatically generated

Figure 1.8 Log Odds versus Quasi-variance statistics of individuals being in School over Training & Apprenticeship

A graph showing the number of numbers and the number of objects

Description automatically generated with medium confidence

Figure 1.9 Log Odds versus Quasi-variance statistics of individuals being in School over Non-traditional Education

### Discussion and Conclusion

The multinomial logistic regression model indicates that structural inequalities do indeed have an impact on an individual’s choice of sorting into economic activity post-mandatory schooling. Educational attainment was the single most significant effect upon individuals entering into employment and unemployment and out of the labour force compared to school. Sex was the single most significant effect upon individuals entering into non-traditional education and training and apprenticeship programs in comparison to school. NS-SEC had a persistent impact on individual activity sorting post-mandatory education. However, this social class impact is less pronounced than educational attainment or sex. Housing tenure plays a small but statistically significant role in all but one of the outcome destinations for post-mandatory schooling youth. The overall conclusion from this model is that structures do matter. However, some structures matter more than others, and this influence changes depending on what type of economic activity is being discussed.

These findings have several implications for previous discussions of social theory. The first relates to the discussion on individualisation and structuralism. These findings present a clear picture that structural inequalities do, in fact, matter. Social class, sex, and housing tenure all present apparent substantive effects on the pathways that individuals choose post-mandatory schooling. Importantly, however, is that different structural inequalities have varied levels of influence and effect sizes dependent upon the given pathway being discussed. For example, some of the most pronounced social class effects relate to individuals choosing to enter employment. Those in NS-SEC 1.1-2 have over a 10 per cent decreased likelihood of being in employment over the school in comparison to their NS-SEC 7 peers. Compared with another structural dimension – such as housing tenure, which only makes up an 8 per cent increased probability of being in employment over school if living in a household that does not own its own home it becomes evident that some structural inequalities matter substantively more than others. Another key example of this relates to the training & apprenticeship pathways. There is a much more pronounced sex-based effect here compared to, say, social class – with men much more likely (24 per cent) to choose this pathway compared to women. Compare this rather substantively significant effect with the impact of being in NS-SEC 1.1 over NS-SEC 7 (4 per cent), and it provides evidence that the influence of certain areas of social stratification depends upon the given type of economic activity. These varied structural effects speak to the complex socio-historical context of the NCDS.

Social class and sex were not the only structural inequalities in this model. Housing tenure was included to assess the views of ‘new structuralism’. Firstly, the view that social class is disaggregated (Saunders, 2003) is demonstrated to be incorrect. That does not necessarily mean that the central tenant of new structuralism –the social restratification of advanced capitalist societies (ibid) presenting new consumption cleavages like housing tenure – is necessarily totally incorrect. Whilst this model presents clear evidence that housing tenure, when controlling for social class, influences an individual’s pathway selection, the effect size across pathways is substantively small. Thus, whilst it would be correct to state that consumption cleavages in the form of housing tenure constitute a level of influence in shaping the material life chances of individuals (Saunders, 2021), there needs to be care not to overstate this influence. Social class matters – above that of housing tenure. Ultimately, whilst new structuralism arguments that social class has become disaggregated do not find empirical evidence amongst the NCDS cohort, the view that housing tenure is essential in influencing pathway choice does find some support. As stated previously, however, the arguments of new structuralism were primarily borne out of the 1980s. The fact that there are any housing tenure effects upon the NCDS cohort is worth noting going forward with comparative analysis in later chapters.

These findings are not without caveats. Firstly, given the socio-historical context of the NCDS cohort, other structural factors, such as race, would have been relevant for inclusion within the model. Unfortunately, practical reasons related to how the NCDS measured race, alongside the weak statistical power of any race variable given the low sample size, mean this is not feasible. Alongside this, combining an unemployment category with an out of the labour force category qualitatively conflates two sociologically distinct concepts – the latter of which has a rich history with structural inequalities related to sex, that for reasons related to statistical power could not be analysed.

In addition, when constructing social class for this model, the choice was made to use NS-SEC. There is no definitive reason for choosing NS-SEC over any other social class schema. The fact that a different choice of schema could potentially influence the substantive interpretation of the findings presented here is cause for concern. Finally, another potential issue relates to missing data. The model presented here covers 8448 individuals. This amounts to 67 per cent of complete cases for sweep 4 (age 23). Missing data could potentially skew the substantive findings. In the next section, this model will undergo a sensitivity analysis alongside other social stratification measures to assess the first issue raised. After this, another section seeks to handle missing data within this model to understand the potential impact missingness has had on the interpretation of this model.

## Sensitivity Analysis of Independent Variables

There are a variety of socio-economic measures used by social scientists. Sensitivity analysis is not common practice within social stratification research (Lambert and Barnett, 2021). However, 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, 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.

### Testing Measures of Parental Social Class

There are strong correlations between parental social class measures. Parental NS-SEC and Parental RGSC have a significant Chi-Square statistic at the p<0.001 level. Parental NS-SEC and CAMSIS have a significant Anova at the p<0.001 level.

Three separate multinomial logistic regressions are presented in Table 1.18. 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.

Focusing first upon the outcome category of employment, NS-SEC and RGSC exhibit identical substantive findings concerning educational attainment, sex, and housing tenure. The CAMSIS model provides identical findings compared to the NS-SEC model for sex but has differences in the substantive findings for both educational attainment and housing tenure – though these differences are minimal (1 per cent and 2 per cent, respectively). The most significant substantive disparity comes from the interpretation of the differing social stratification variables themselves. Whilst NS-SEC and RGSC all show a general pattern that characterises an increased log odds of being in employment over education as NS-SEC or RGSC increases from the reference category, the substantive interpretation from these models is different. CAMSIS as a model does not present any substantive findings related to CAMSIS effects on economic activity sorting. Translated into average marginal effects, this presents a 0 per cent increased probability of being employed over the school for each unit increase in CAMSIS. This is a substantively different finding to the NS-SEC model that states there is anywhere from a 5 per cent to 13 per cent increased probability of being in employment over school compared to NS-SEC 2 peers. The NS-SEC and RGSC models, on the other hand, whilst differing on exact numbers, provide the same substantive conclusions to one another. There is a general trend that as you go down the class schema, there is an increased probability of being in employment over school compared to their reference category peers.

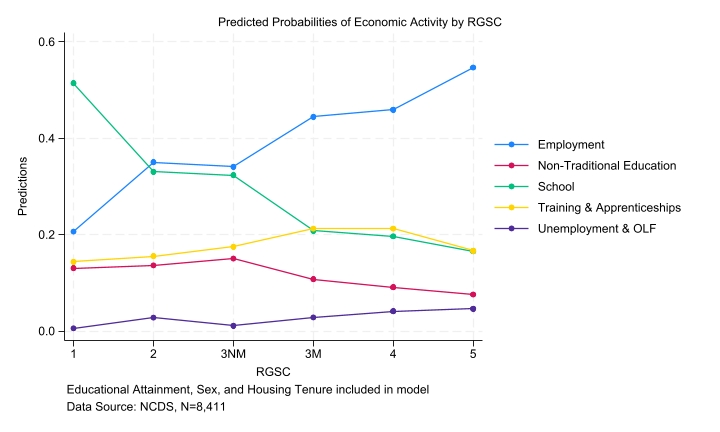
Moving on to the non-traditional education category there are identical findings for sex and housing tenure across all three models. With the exception of a 1 per cent difference within the RGSC model, the same can be said for educational attainments substantive interpretation across all three models. There are also substantively difference interpretations between all three models. The CAMSIS model presents a statistically significant result though it is non-substantively significant. NS-SEC compared to RGSC finds that those at the bottom of the social stratification schema have a lower log odds of being in non-traditional education over school compared to their reference category – this results in a 3 per cent decreased probability. This is not found within the RGSC model.

Moving on to the training & apprenticeship category there are identical findings for educational attainment, sex, and housing tenure across all three models. There are also substantively identical interpretations to be drawn from the NS-SEC and RGSC models – whilst CAMSIS again reports a statistically significant, albeit non-substantively significant result.

Finally, the unemployment & OLF category sees identical findings for educational attainment, sex, and housing tenure across all three models. Whilst once again, the CAMSIS model finds no substantive impact that social stratification has upon individuals sorting into economic activity post-mandatory schooling, both NS-SEC and RGSC models do. This substantive significance however is minor – the largest impact from NS-SEC reports a 3 per cent decreased probability for NS-SEC 1.2 social origins compared to the reference category of being in unemployment & OLF over school. For the RGSC model the largest impact is even smaller at 2 per cent – individuals from RGSC 1 compared to the reference category had a 2 per cent decreased probability of being in unemployment & OLF over school. For a more intuitive comparison of the substantive similarities and differences between RGSC and NS-SEC the predicted probabilities are graphed below in figures XXXX and XXXX.

A graph showing the number of people in the united states

Description automatically generated



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 NS-SEC model when using AIC and the CAMSIS model when using 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 schema. These differences are, however, minimal. Given the minor substantive differences found within the two models (mainly located within the employment category) alongside the slight preference for NS-SEC using AIC, there is a general preference towards the NS-SEC model. As such, going forward, the preferred model of choice for subsequent analysis will be the NS-SEC model.

Table 2.18 Sensitivity analyses of alternative measures of parental social stratification

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NS-SEC | | | | | CAMSIS | | | | | RGSC | | | | |
| Economic Activity | **Coef.** | **S.E** | **Sig.** | **Prob.** | **S.E** | **Coef.** | **S.E** | **Sig.** | **Prob.** | **S.E** | **Coef.** | **S.E** | **Sig.** | **Prob.** | **S.E** |
| Employment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.58 | (0.08) | \*\*\* | -0.39 | (0.01) | -3.57 | (0.08) | \*\*\* | -0.38 | (0.01) | -3.60 | (0.08) | \*\*\* | -0.39 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) |  |  |  | (.) | (.) |  |  |  | (.) | (.) |
| *Male* | -0.40 | (0.07) | \*\*\* | -0.17 | (0.01) | -0.39 | (0.07) | \*\*\* | -0.17 | (0.01) | -0.40 | (0.07) | \*\*\* | -0.17 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) |  |  |  | (.) | (.) |  |  |  | (.) | (.) |
| *Do not Own Home* | 0.68 | (0.08) | \*\*\* | 0.08 | (0.01) | 0.65 | (0.08) | \*\*\* | 0.08 | (0.01) | 0.71 | (0.08) | \*\*\* | 0.08 | (0.01) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1.1* | 0.08 | (0.23) |  | 0.02 | (0.03) |  |  |  |  |  |  |  |  |  |  |
| *1.2* | -0.67 | (0.22) | \*\* | -0.05 | (0.03) |  |  |  |  |  |  |  |  |  |  |
| *2* | Ref. | (.) |  | (.) | (.) |  |  |  |  |  |  |  |  |  |  |
| *3* | 0.36 | (0.15) | \* | 0.05 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| *4* | 0.96 | (0.14) | \*\*\* | 0.08 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| *5* | 0.87 | (0.14) | \*\*\* | 0.07 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| *6* | 0.91 | (0.14) | \*\*\* | 0.09 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| *7* | 1.36 | (0.13) | \*\*\* | 0.13 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| CAMSIS (SOC 2000) |  |  |  |  |  | -0.04 | (0.00) | \*\*\* | -0.00 | (0.00) |  |  |  |  |  |
| RGSC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1* |  |  |  |  |  |  |  |  |  |  | -0.97 | (0.23) | \*\*\* | -0.10 | (0.03) |
| *2* |  |  |  |  |  |  |  |  |  |  | Ref. | (.) |  | (.) | (.) |
| *3NM* |  |  |  |  |  |  |  |  |  |  | -0.01 | (0.13) |  | -0.01 | (0.02) |
| *3M* |  |  |  |  |  |  |  |  |  |  | 0.70 | (0.10) | \*\*\* | 0.05 | (0.01) |
| *4* |  |  |  |  |  |  |  |  |  |  | 0.79 | (0.13) | \*\*\* | 0.06 | (0.02) |
| *5* |  |  |  |  |  |  |  |  |  |  | 1.14 | (0.16) | \*\*\* | 0.12 | (0.02) |
| Intercept | 0.89 | (0.12) | \*\*\* |  |  | 3.45 | (0.16) | \*\*\* |  |  | 1.17 | (0.10) | \*\*\* |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Non-Traditional Education |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -1.71 | (0.09) | \*\*\* | 0.03 | (0.01) | -1.70 | (0.09) | \*\*\* | 0.02 | (0.01) | -1.71 | (0.09) | \*\*\* | 0.03 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | -0.84 | (0.09) | \*\*\* | -0.07 | (0.01) | -0.83 | (0.09) | \*\*\* | -0.07 | -0.07 | -0.84 | (0.09) | \*\*\* | -0.07 | -0.07 |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | -0.11 | (0.10) |  | -0.04 | (0.01) | -0.12 | (0.09) |  | -0.04 | (0.01) | -0.08 | (0.10) |  | -0.04 | (0.01) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1.1* | 0.02 | (0.23) |  | 0.00 | (0.02) | (.) | (.) |  |  |  | (.) | (.) |  |  |  |
| *1.2* | -0.52 | (0.21) | \* | -0.01 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| *2* | Ref. | (.) |  | (.) | (.) |  |  |  |  |  |  |  |  |  |  |
| *3* | 0.08 | (0.16) |  | -0.01 | (0.01) |  |  |  |  |  |  |  |  |  |  |
| *4* | 0.15 | (0.17) |  | -0.03 | (0.01) |  |  |  |  |  |  |  |  |  |  |
| *5* | 0.24 | (0.15) |  | -0.02 | (0.01) |  |  |  |  |  |  |  |  |  |  |
| *6* | -0.02 | (0.16) |  | -0.04 | (0.01) |  |  |  |  |  |  |  |  |  |  |
| *7* | 0.44 | (0.15) | \*\* | -0.03 | (0.01) |  |  |  |  |  |  |  |  |  |  |
| CAMSIS (SOC 2000) |  |  |  |  |  | -0.01 | (0.00) | \*\*\* | 0.00 | (0.00) |  |  |  |  |  |
| RGSC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1* |  |  |  |  |  |  |  |  |  |  | -0.49 | (0.21) | \* | 0.00 | (0.02) |
| *2* |  |  |  |  |  |  |  |  |  |  | Ref. | (.) |  | (.) | (.) |
| *3NM* |  |  |  |  |  |  |  |  |  |  | 0.12 | (0.14) |  | 0.01 | (0.01) |
| *3M* |  |  |  |  |  |  |  |  |  |  | 0.23 | (0.11) | \* | -0.02 | (0.01) |
| *4* |  |  |  |  |  |  |  |  |  |  | 0.11 | (0.16) |  | -0.03 | (0.01) |
| *5* |  |  |  |  |  |  |  |  |  |  | 0.10 | (0.23) |  | -0.04 | (0.01) |
| Intercept | 0.18 | (0.13) |  |  |  | 0.84 | (0.19) | \*\*\* | (.) | (.) | 0.18 | (0.11) |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Training & Apprenticeships |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.24 | (0.09) | \*\*\* | -0.17 | (0.01) | -3.21 | (0.09) | \*\*\* | -0.17 | (0.01) | -3.25 | (0.09) | \*\*\* | -0.17 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 1.52 | (0.09) | \*\*\* | 0.24 | (0.01) | 1.52 | (0.09) | \*\*\* | 0.24 | (0.01) | 1.52 | (0.09) | \*\*\* | 0.24 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.38 | (0.08) | \*\*\* | -0.01 | (0.01) | 0.34 | (0.08) | \*\*\* | -0.01 | (0.01) | 0.40 | (0.08) | \*\*\* | -0.01 | (0.01) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1.1* | -0.12 | (0.26) |  | -0.02 | (0.03) | (.) | (.) |  |  |  | (.) | (.) |  |  |  |
| *1.2* | -0.45 | (0.23) |  | 0.00 | (0.03) |  |  |  |  |  |  |  |  |  |  |
| *2* | Ref. | (.) |  | (.) | (.) |  |  |  |  |  |  |  |  |  |  |
| *3* | 0.19 | (0.17) |  | -0.00 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| *4* | 0.97 | (0.16) | \*\*\* | 0.05 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| *5* | 0.90 | (0.15) | \*\*\* | 0.05 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| *6* | 0.84 | (0.15) | \*\*\* | 0.04 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| *7* | 1.08 | (0.15) | \*\*\* | 0.02 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| CAMSIS (SOC 2000) |  |  |  |  |  | -0.04 | (0.00) | \*\*\* | -0.00 | (0.00) |  |  |  |  |  |
| RGSC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1* |  |  |  |  |  | (.) | (.) |  |  |  | -0.52 | (0.23) | \* | 0.01 | (0.03) |
| *2* |  |  |  |  |  |  |  |  |  |  | Ref. | (.) |  | (.) | (.) |
| *3NM* |  |  |  |  |  |  |  |  |  |  | 0.14 | (0.15) |  | 0.02 | (0.02) |
| *3M* |  |  |  |  |  |  |  |  |  |  | 0.78 | (0.11) | \*\*\* | 0.04 | (0.01) |
| *4* |  |  |  |  |  |  |  |  |  |  | 0.83 | (0.14) | \*\*\* | 0.04 | (0.01) |
| *5* |  |  |  |  |  |  |  |  |  |  | 0.76 | (0.19) | \*\*\* | -0.00 | (0.02) |
| Intercept | -0.72 | (0.14) | \*\*\* |  |  | 1.53 | (0.18) | \*\*\* |  |  | -0.56 | (0.12) | \*\*\* |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Unemployment & Out of Labour Force |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.94 | (0.27) | \*\*\* | -0.03 | (0.00) | -3.93 | (0.27) | \*\*\* | -0.03 | (0.00) | -3.97 | (0.27) | \*\*\* | -0.03 | (0.00) |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | -0.37 | (0.14) | \*\* | -0.01 | (0.00) | -0.36 | (0.14) | \*\* | -0.01 | (0.00) | -0.37 | (0.14) | \* | -0.01 | (0.00) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.8 | (0.16) | \*\*\* | 0.01 | (0.00) | 0.81 | (0.15) | \*\*\* | 0.01 | (0.00) | 0.91 | (0.16) | \*\*\* | 0.01 | (0.00) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1.1* | -0.35 | (0.67) |  | -0.01 | (0.02) | (.) | (.) |  |  |  | (.) | (.) |  |  |  |
| *1.2* | -2.08 | (1.03) | \* | -0.03 | (0.01) |  |  |  |  |  |  |  |  |  |  |
| *2* | Ref. | (.) |  | (.) | (.) |  |  |  |  |  |  |  |  |  |  |
| *3* | -0.29 | (0.38) |  | -0.01 | (0.01) |  |  |  |  |  |  |  |  |  |  |
| *4* | 0.60 | (0.31) |  | -0.00 | (0.01) |  |  |  |  |  |  |  |  |  |  |
| *5* | 0.29 | (0.30) |  | -0.01 | (0.01) |  |  |  |  |  |  |  |  |  |  |
| *6* | 0.77 | (0.28) | \*\* | 0.00 | (0.01) |  |  |  |  |  |  |  |  |  |  |
| *7* | 1.20 | (0.27) | \*\*\* | 0.01 | (0.01) |  |  |  |  |  |  |  |  |  |  |
| CAMSIS (SOC 2000) |  |  |  |  |  | -0.05 | (0.01) | \*\*\* | -0.00 | (0.00) |  |  |  |  |  |
| RGSC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1* |  |  |  |  |  | (.) | (.) |  |  |  | -2.02 | (1.03) | \* | -0.02 | (0.01) |
| *2* |  |  |  |  |  |  |  |  |  |  | Ref. | (.) |  | (.) | (.) |
| *3NM* |  |  |  |  |  |  |  |  |  |  | -0.91 | (0.35) | \* | -0.02 | (0.01) |
| *3M* |  |  |  |  |  |  |  |  |  |  | 0.48 | (0.21) | \* | -0.00 | (0.01) |
| *4* |  |  |  |  |  |  |  |  |  |  | 0.91 | (0.24) | \*\*\* | 0.01 | (0.01) |
| *5* |  |  |  |  |  |  |  |  |  |  | 1.19 | (0.28) | \*\*\* | 0.01 | (0.01) |
| Intercept | -1.48 | (0.25) | \*\*\* |  |  | 1.35 | (0.33) | \*\*\* |  |  | -1.35 | (0.21) | \*\*\* |  |  |
| Number of observations | 8411 | | | | | 8411 | | | | | 8411 | | | | |
| McFadden’s | 0.25 | | | | | 0.25 | | | | | 0.25 | | | | |
| McFadden’s Adjusted Pseudo | 0.24 | | | | | 0.24 | | | | | 0.24 | | | | |
| Cox-Snell Pseudo | 0.49 | | | | | 0.49 | | | | | 0.49 | | | | |
| Nagelkerke Pseudo | 0.53 | | | | | 0.52 | | | | | 0.52 | | | | |
| AIC | 17431.50 | | | | | 17414.46 | | | | | 17454.71 | | | | |
| BIC | 17741.14 | | | | | 17555.21 | | | | | 17708.05 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: Complete Records Analysis for NS-SEC, CAMSIS, RGSC | | | | | | | | | | | | | | | |

### Discussion and Conclusions

This section has sought to provide a sensitivity analysis of socio-economic measures into the model of economic activity to understand if there is any temporal sensitivity amongst the socio-economic measures and also to understand the best-fit model amongst these measures. Overall, the findings suggest that there are some differences between social stratification measures within each model – particularly concerning the employment category of economic activity. Beyond this, however, there are relatively consistent, stable substantive findings between the models – though CAMSIS, out of the three models, is most likely to have diverging findings.

Perhaps the most exciting finding of this sensitivity analysis stems from the near-identical results from the NS-SEC and RGSC models. The NS-SEC predecessor – the Goldthorpe schema – was claimed by the likes of Marshal to be superior to the likes of the RGSC schema (Rose and Pevalin, 2002). For the sensitivity analysis to present substantive findings of the Goldthorpe schemas successor being identical to the RGSC does present some challenges to the apparent superiority of this analytical construction.

When comparing AIC statistics, there is a slight advantage for the NS-SEC model. When comparing BIC statistics, there is a slight advantage for the NS-SEC model also. The NS-SEC model is selected going forward.

Overall, this section on sensitivity analysis has used contemporary statistical techniques to update prior literature on youth transitions within the NCDS cohort and has also aided in model selection going forward. The following section seeks to continue this tradition of employing contemporary statistical techniques by attempting to deal with missingness within the NS-SEC preferred model.

## Sensitivity analysis using SOC codes

Given the sensitivity analysis of social stratification measures, NS-SEC has been found to be the best measure to use within the model. Following this analysis, 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 is found within the appendix.

### SOC codes Modelling

The following tables – 1.19 and 1.20 follow a similar design to those tables 1.15 and 1.16 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 1.19, whilst NS-SEC using the SOC 2000 construction as a difference of 1268.65 deviance from the null, the SOC 90 construction has 1009.70 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 1.20, the full model with SOC 2000 construction of NS-SEC has a 259.42 difference in deviance from the previous model compared to a 190.99 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 2.19 Goodness-of-fit summaries for explanatory variables and Economic Activity Comparing SOC codes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 23040.88 | - | - | - | 23048.88 | 23077.03 |
| Null Model + Educational Attainment | 18829.53 | 4211.35 | 4 | 0.18 | 18845.83 | 18901.83 |
| Null Model + Sex | 22052.16 | 988.72 | 4 | 0.04 | 22068.16 | 22124.46 |
| Null Model + Tenure | 22224.78 | 816.11 | 4 | 0.04 | 22240.77 | 22297.07 |
| Null Model + NS-SEC (SOC 2000) | 21772.23 | 1268.65 | 28 | 0.06 | 21836.23 | 22061.42 |
| Null Model + NS-SEC (SOC 90) | 22031.18 | 1009.70 | 28 | 0.04 | 22095.18 | 22320.38 |

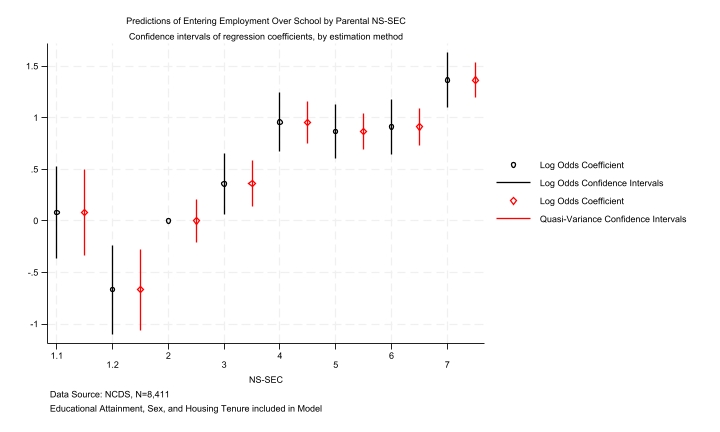
Explanatory variables are entered sequentially in the subsequent multiple logistic model following the (Gayle and Lambert, 2009) example.

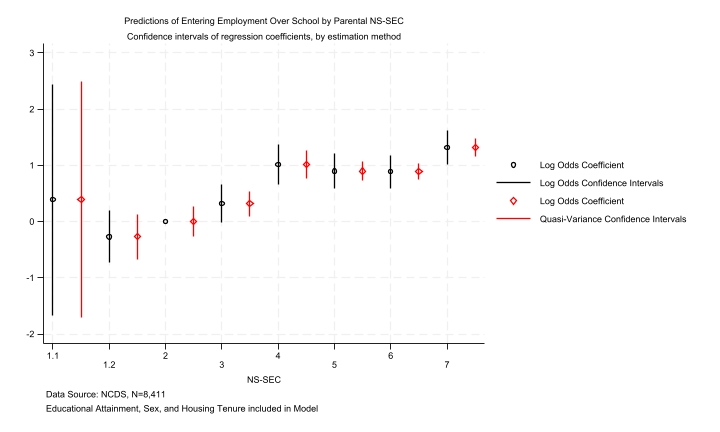
Table 2.20 Model building goodness-of-fit summaries for multinominal logistic regression model of Economic Activity Comparing SOC codes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 23040.88 | - | - | - | 23048.88 | 23077.03 |
| Null Model + Educational Attainment | 18829.53 | 4211.35 | 4 | 0.18 | 18845.83 | 18901.83 |
| Null Model + Educational Attainment + Sex | 17843.77 | 985.76 | 4 | 0.23 | 17867.77 | 17952.22 |
| Null Model + Educational Attainment + Sex + Tenure | 17602.92 | 240.85 | 4 | 0.24 | 17634.92 | 17747.51 |
| Null Model + Educational Attainment + Sex + Tenure + NS-SEC (SOC 2000) | 17343.50 | 259.42 | 28 | 0.25 | 17431.50 | 17741.14 |
| Null Model + Educational Attainment + Sex + Tenure + NS-SEC (SOC 90) | 17411.93 | 190.99 | 28 | 0.24 | 17499.93 | 17809.57 |

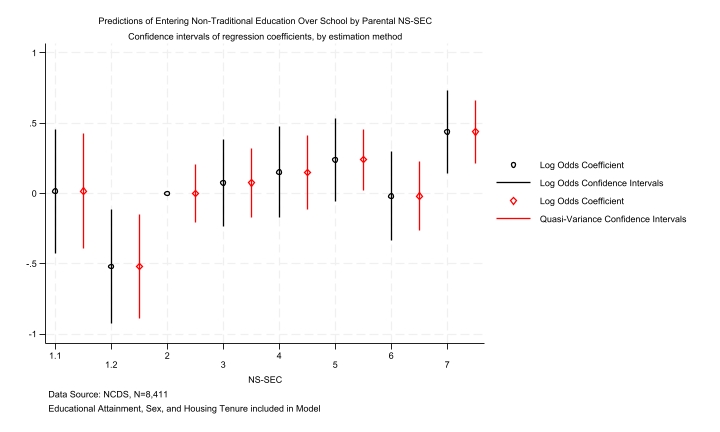
Table 1.21 displays a comparison of the full proposed model using the SOC 2000 construction of NS-SEC in one model and the SOC 90 construction of NS-SEC in the second model. Both log odds coefficients and average marginal effects statistics are provided for ease of interpretation. Average marginal effects statistics were included to assess the degree of difference or similarity between the two models as log odds is a difficult statistic to measure and understand even small differences within it. Unsurprisingly educational attainment, sex, and housing tenure do not deviate beyond a single incident across the two models[[11]](#footnote-11).

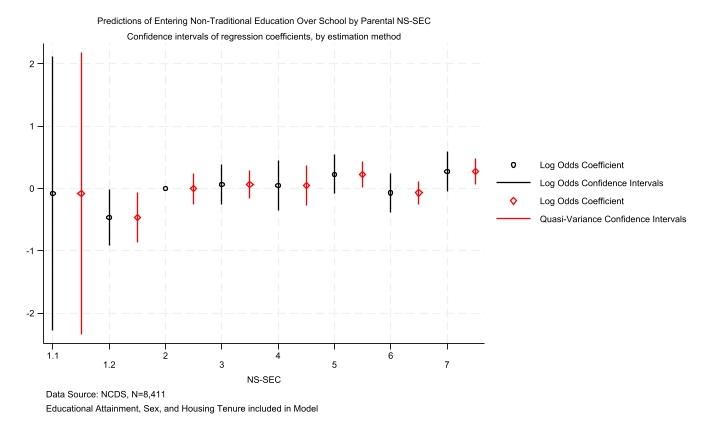
Starting with the employment category, whilst NS-SEC 1.2 and 3 are statistically significant within the SOC 2000 model this is not the case for the SOC 90 model. Both models share a statistical significance for NS-SEC 4-7. The general trend from these categories is that the SOC 90 model provides a larger effect size across NS-SEC 4-7 in comparison to the SOC 2000 model – though both models present the same general trend of an increased probability of being in employment over school in comparison to NS-SEC 2 as you go down the class schema. Each models log odds and quasi-variance statistics are graphed in figures XXXX and XXXX.



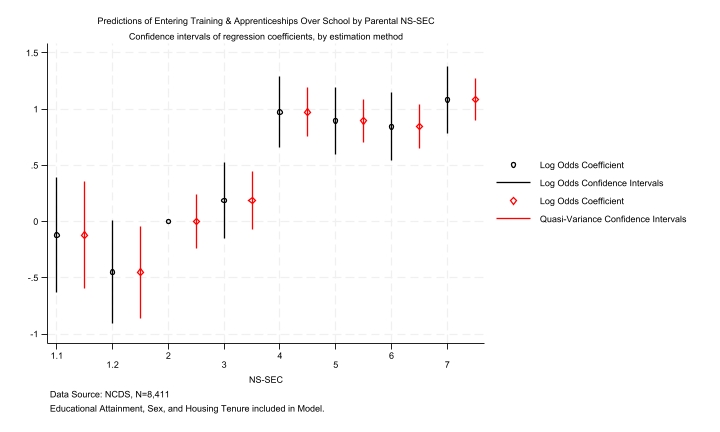


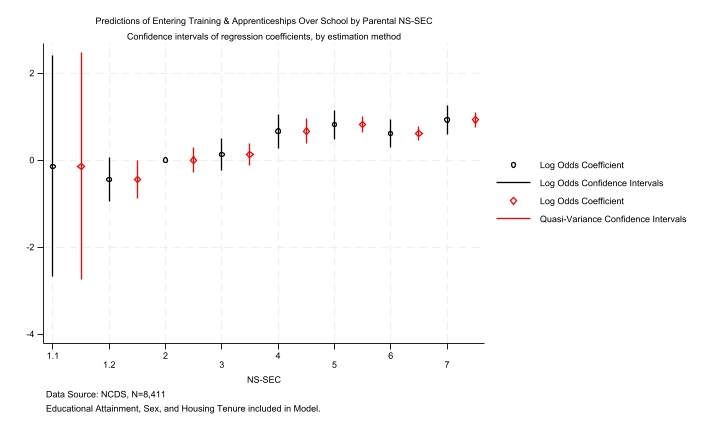
Moving on to the non-traditional education category and once more there is an instance of an NS-SEC category, this time NS-SEC 7 being statistically significant in the SOC 2000 model but not in the SOC 90 model. Both models find NS-SEC 1.2 to be statistically significant. Once again, the SOC 90 model agrees with the general substantive trend of the SOC 2000 model – in this instance a decreased probability of being in non-traditional education over school compared to NS-SEC 2 for individuals in NS-SEC 1.2. However, SOC 90 states this effect size at 2 per cent decreased probability to SOC 2000s 1 per cent decreased probability. Each models log odds and quasi-variance statistics are graphed in figures XXXX and XXXX.





Both the SOC 2000 model and the SOC 90 model find NS-SEC 4-7 to be statistically significant for the training & apprenticeship category. SOC 90 in comparison to the SOC 2000 model reduces the effect sizes of all statistically significant NS-SEC categories for the training & apprenticeship category. The general trend, once again, is similar across the models but this time the SOC 90 model reduces the effect size. Each models log odds and quasi-variance statistics are graphed in figures XXXX and XXXX.





Finally, with respect to the unemployment & OLF category, the SOC 2000 model finds NS-SEC 1.2 and 6 to be statistically significant when the SOC 90 model does not. Both however find NS-SEC 7 statistically significant. Both agree on both the trend and substantive effect of this category at 1 per cent increased probability of being in unemployment & OLF over school compared to NS-SEC 2. The log odds and quasi-variance could not be graphically compared for the unemployment & OLF category because of the low number of observations in the SOC 90 category making the standard errors too large to compare across models in a graph.

The primary reason for this sensitivity analysis was a direct comparison of NS-SEC under two different constructions using SOC 90 and SOC 2000 codes. For this reason, a graphical comparison of the predicted probabilities of both measures is also provided to see a more intuitive direct comparison between the two measures. Figures XXXX and XXXX graph the predicted probabilities of NS-SEC using both SOC 90 and SOC 2000 codes.

A graph showing the number of people in the united states

Description automatically generated

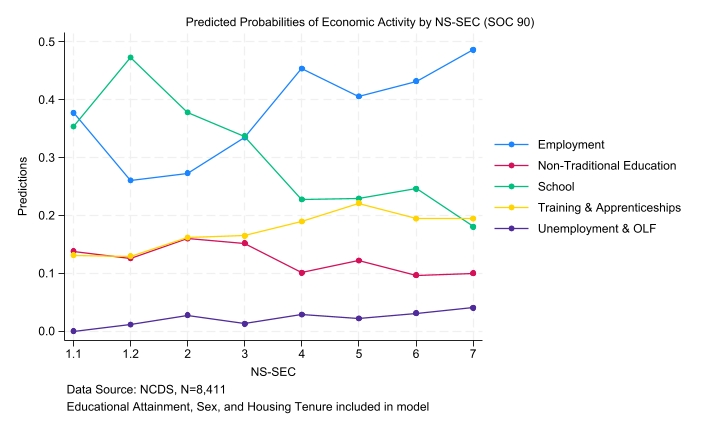


Table 2.21 Sensitivity analyses of SOC codes

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SOC 2000 Codes | | | | | SOC 90 Codes | | | | | |
|  | NS-SEC | | | Average Marginal Effects | | NS-SEC | | | | Average Marginal Effects | |
| Economic Activity: ‘School’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **Coef.** | **S.E.** | **Sig.** | | **Prob.** | **S.E.** |
| Employment |  |  |  |  |  |  |  |  | |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  | |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *Five or More O’levels* | -3.58 | (0.08) | \*\*\* | -0.39 | (0.01) | -3.61 | (0.08) | \*\*\* | | -0.39 | (0.01) |
| Sex |  |  |  |  |  |  |  |  | |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *Male* | -0.40 | (0.07) | \*\*\* | -0.17 | (0.01) | -0.41 | (0.07) | \*\*\* | | -0.17 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  | |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *Do not Own Home* | 0.68 | (0.08) | \*\*\* | 0.08 | (0.01) | 0.74 | (0.08) | \*\*\* | | 0.09 | (0.01) |
| NS-SEC |  |  |  |  |  |  |  |  | |  |  |
| *1.1* | 0.08 | (0.23) |  | 0.02 | (0.03) | 0.39 | (1.05) |  | | 0.09 | (0.17) |
| *1.2* | -0.67 | (0.22) | \*\* | -0.05 | (0.03) | -0.27 | (0.24) |  | | 0.01 | (0.04) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *3* | 0.36 | (0.15) | \* | 0.05 | (0.02) | 0.32 | (0.17) |  | | 0.05 | (0.02) |
| *4* | 0.96 | (0.14) | \*\*\* | 0.08 | (0.02) | 1.02 | (0.18) | \*\*\* | | 0.12 | (0.03) |
| *5* | 0.87 | (0.14) | \*\*\* | 0.07 | (0.02) | 0.90 | (0.16) | \*\*\* | | 0.08 | (0.02) |
| *6* | 0.91 | (0.14) | \*\*\* | 0.09 | (0.02) | 0.89 | (0.15) | \*\*\* | | 0.10 | (0.02) |
| *7* | 1.36 | (0.13) | \*\*\* | 0.13 | (0.02) | 1.32 | (0.15) | \*\*\* | | 0.14 | (0.02) |
| Intercept | 0.89 | (0.12) | \*\*\* | (.) | (.) | 0.78 | (0.14) | \*\*\* | | (.) | (.) |
|  |  |  |  |  |  |  |  |  | |  |  |
| Non-Traditional Education |  |  |  |  |  |  |  |  | |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  | |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *Five or More O’levels* | -1.71 | (0.09) | \*\*\* | 0.03 | (0.01) | -1.72 | (0.09) | \*\*\* | | 0.03 | (0.01) |
| Sex |  |  |  |  |  |  |  |  | |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *Male* | -0.84 | (0.09) | \*\*\* | -0.07 | (0.01) | -0.84 | (0.09) | \*\*\* | | -0.07 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  | |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *Do not Own Home* | -0.11 | (0.10) |  | -0.04 | (0.01) | -0.07 | (0.10) |  | | -0.04 | (0.01) |
| NS-SEC |  |  |  |  |  |  |  |  | |  |  |
| *1.1* | 0.02 | (0.23) |  | 0.00 | (0.02) | -0.08 | (1.12) |  | | -0.02 | (0.10) |
| *1.2* | -0.52 | (0.21) | \* | -0.01 | (0.02) | -0.46 | (0.23) | \* | | -0.02 | (0.02) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *3* | 0.08 | (0.16) |  | -0.01 | (0.01) | 0.06 | (0.17) |  | | -0.01 | (0.02) |
| *4* | 0.15 | (0.17) |  | -0.03 | (0.01) | 0.05 | (0.20) |  | | -0.04 | (0.02) |
| *5* | 0.24 | (0.15) |  | -0.02 | (0.01) | 0.23 | (0.16) |  | | -0.02 | (0.01) |
| *6* | -0.02 | (0.16) |  | -0.04 | (0.01) | -0.07 | (0.16) |  | | -0.05 | (0.01) |
| *7* | 0.44 | (0.15) | \*\* | -0.03 | (0.01) | 0.27 | (0.16) |  | | -0.04 | (0.01) |
| Intercept | 0.18 | (0.13) |  | (.) | (.) | 0.20 | (0.14) |  | | (.) | (.) |
|  |  |  |  |  |  |  |  |  | |  |  |
| School | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
|  |  |  |  |  |  |  |  |  | |  |  |
| Training & Apprenticeships |  |  |  |  |  |  |  |  | |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  | |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *Five or More O’levels* | -3.24 | (0.09) | \*\*\* | -0.17 | (0.01) | -3.26 | (0.09) | \*\*\* | | -0.17 | (0.01) |
| Sex |  |  |  |  |  |  |  |  | |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *Male* | 1.52 | (0.09) | \*\*\* | 0.24 | (0.01) | 1.51 | (0.09) | \*\*\* | | 0.24 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  | |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *Do not Own Home* | 0.38 | (0.08) | \*\*\* | -0.01 | (0.01) | 0.45 | (0.08) | \*\*\* | | -0.01 | (0.01) |
| NS-SEC |  |  |  |  |  |  |  |  | |  |  |
| *1.1* | -0.12 | (0.26) |  | -0.02 | (0.03) | -0.14 | (1.30) |  | | -0.03 | (0.13) |
| *1.2* | -0.45 | (0.23) |  | 0.00 | (0.03) | -0.45 | (0.26) |  | | -0.02 | (0.03) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *3* | 0.19 | (0.17) |  | -0.00 | (0.02) | 0.13 | (0.19) |  | | -0.00 | (0.02) |
| *4* | 0.97 | (0.16) | \*\*\* | 0.05 | (0.02) | 0.67 | (0.20) | \*\*\* | | 0.01 | (0.02) |
| *5* | 0.90 | (0.15) | \*\*\* | 0.05 | (0.02) | 0.81 | (0.17) | \*\*\* | | 0.04 | (0.02) |
| *6* | 0.84 | (0.15) | \*\*\* | 0.04 | (0.02) | 0.61 | (0.16) | \*\*\* | | 0.01 | (0.02) |
| *7* | 1.08 | (0.15) | \*\*\* | 0.02 | (0.02) | 0.93 | (0.17) | \*\*\* | | 0.01 | (0.02) |
| Intercept | -0.72 | (0.14) | \*\*\* | (.) | (.) | -0.68 | (0.16) | \*\*\* | | (.) | (.) |
|  |  |  |  |  |  |  |  |  | |  |  |
| Unemployment & Out of Labour Force |  |  |  |  |  |  |  |  | |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  | |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *Five or More O’levels* | -3.94 | (0.27) | \*\*\* | -0.03 | (0.00) | -3.97 | (0.27) | \*\*\* | | -0.03 | (0.00) |
| Sex |  |  |  |  |  |  |  |  | |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *Male* | -0.37 | (0.14) | \*\* | -0.01 | (0.00) | -0.38 | (0.14) | \*\* | | -0.01 | (0.00) |
| Housing Tenure |  |  |  |  |  |  |  |  | |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *Do not Own Home* | 0.8 | (0.16) | \*\*\* | 0.01 | (0.00) | 0.92 | (0.16) | \*\*\* | | 0.01 | (0.00) |
| NS-SEC |  |  |  |  |  |  |  |  | |  |  |
| *1.1* | -0.35 | (0.67) |  | -0.01 | (0.02) | -11.01 | (650.75) |  | | -0.03 | (0.01) |
| *1.2* | -2.08 | (1.03) | \* | -0.03 | (0.01) | -1.13 | (0.78) |  | | -0.02 | (0.01) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | | (.) | (.) |
| *3* | -0.29 | (0.38) |  | -0.01 | (0.01) | -0.69 | (0.47) |  | | -0.02 | (0.01) |
| *4* | 0.60 | (0.31) |  | -0.00 | (0.01) | 0.54 | (0.39) |  | | -0.00 | (0.01) |
| *5* | 0.29 | (0.30) |  | -0.01 | (0.01) | 0.27 | (0.35) |  | | -0.01 | (0.01) |
| *6* | 0.77 | (0.28) | \*\* | 0.00 | (0.01) | 0.55 | (0.33) |  | | -0.00 | (0.01) |
| *7* | 1.20 | (0.27) | \*\*\* | 0.01 | (0.01) | 1.11 | (0.32) | \*\*\* | | 0.01 | (0.01) |
| Intercept | -1.48 | (0.25) | \*\*\* | (.) | (.) | -1.50 | (0.31) | \*\*\* | | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Number of observations | 8411 | | | | | 8411 | | | | | |
| McFadden’s | 0.25 | | | | | 0.24 | | | | | |
| McFadden’s Adjusted Pseudo | 0.24 | | | | | 0.24 | | | | | |
| Cox-Snell Pseudo | 0.49 | | | | | 0.49 | | | | | |
| Nagelkerke Pseudo | 0.53 | | | | | 0.52 | | | | | |
| AIC | 17431.50 | | | | | 17499.93 | | | | | |
| BIC | 17741.14 | | | | | 17809.57 | | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: Complete Records Analysis | | | | | | | | | | | |

### Discussion and Conclusion

This section was an attempt to understand the possible differences and similarities between the same analytical measure with different SOC code constructions. A sensitivity analysis was conducted using a SOC 2000 construction of NS-SEC and a SOC 90 construction. Given that the NCDS dataset began in 1958, intuitively, the original prior supposed that a SOC 90 construction of any social stratification measure would more accurately capture the social structures of the NCDS cohort compared to the SOC 2000 counterpart. Goodness-of-fit statistics have however presented an alternative narrative. Both AIC and BIC statistics favour the SOC 2000 model – though the overall difference is very small. Whilst somewhat surprising, these results do allow the prior discussion on the advantages and disadvantages of using SOC 90 constructions to be updated. It appears that the SOC 90 construction of NS-SEC does not capture the social structure of the NCDS cohort better than the SOC 2000 construction. Given this, the SOC 2000 model is the preferred model going forward. Whilst the substantive results across both models followed the same general positive/negative trends there were distinct discrepancies on the size of these substantive effects particularly for the impact of NS-SEC on the model. For employment and non-traditional education categories SOC 90 appeared to inflate the influence of the substantive effects of NS-SEC in comparison to the SOC 2000 model. For training & apprenticeship and unemployment & OLF categories, this was reversed, and the SOC 90 model appeared to diminish the influence of the substantive effects of NS-SEC in comparison to the SOC 2000 model. In particular, if the SOC 90 model would have been produced without a sensitivity analysis of SOC codes, the training & apprenticeship category would imply that whilst there is a statistically significant effect across many NS-SEC categories, the substantive effect is near zero. The SOC 2000 model instead has a difference of up to 4 per cent at times, ultimately changing the overall substantive picture within the training & apprenticeship category. Overall, from a statistical as well as substantive viewpoint, this sensitivity analysis was an effective exercise.

## Missing Data in the NCDS

### 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 above 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 – utilising this partial data within my analysis could be beneficial.

When dealing with missing data, there are three primary types of classification. The first is missing completely at random (MCAR), meaning missingness does not depend on observed or unobserved values. The second is being missed at random (MAR), meaning that given observed values, missingness does not depend on the unobserved ones. Finally, missing not at random (MNAR) means missingness depends on unobserved values (Silverwood et al. 2021). If data is found to be MAR, then approaches like multiple imputation (MI) and inverse probability weighting are made available – the former being extensively documented with the NCDS in particular (Hawkes and Plewis 2006).

When dealing with missing data, there are multiple methods to tackle the problem. 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 (Carpenter and Kenward, 2012).

A second method that deals with missing data is the use of survey weights. Survey weights take into account 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 (Seaman *et al.*, 2012; Seaman and White, 2013).

A third method involves Multiple Imputation (MI). This method substituted missing data with substituted values[[12]](#footnote-12). MI is an attractive method because it is practical and widely applicable (Carpenter and Kenward, 2012).

Full-information Maximum Likelihood (FIML) is another method for dealing with missingness. For the regression-based analysis, including interactions with data from at least two stages of the life course (Silverwood et al., 2021), as the current analysis is, multiple imputation is plausible and more flexible than FIML. This flexibility stems from the ability to include auxiliary variables more easily within the imputation phase as well as being readily able to, after imputing data sets, obtain point estimates and standard errors at ease (Carpenter and Kenward, 2012). Recently, there has been some debate surrounding FIML vs MI approaches.

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[[13]](#footnote-13). 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[[14]](#footnote-14). Overall, whilst FIML does offer some advantages, there is nothing so considerable 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 Study

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 a random normal distribution of 1000 observations around a normally distributed metric dependent variable and three independent variables that share an identical distribution. Each independent variable has the same level of correlation associated with the dependent variable. This is to again allow for a point of comparison when missingness is injected into one variable and not the others to see what happens when handling missing data practices are implemented. 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.

For the first set of models in this simulation missingness is introduced using an MCAR mechanism to check that the simulation is working correctly and that all the ‘good’ ways to handle missing data return a near identical result to the ‘God’ models. Results are presented in table XXXX. Missingness is injected into independent variable three **[include how much].** This missingness accounts for 49 per cent missingness in the model. Because this missingness is MCAR the results are identical to the previous two models. Next, we move on to our ‘bad’ handling missing data practices. Independent variable three is a dummy binary variable, first a model is produced whereby all missingness is coded as =0 and another is produced =1, both models perform poorly and skew the substantive results considerably. Next a single use model imputation is used – the same framework as a single use mean imputation but because the variable is categorical mode is used over mean. This model performs well and it should, given that this is an MCAR mechanism. Finally, the next set of models are all under the ‘good’ handling missing data practices. First is the FIML model that presents identical results to the ‘god’ model. Now we move to the MI approaches, three are displayed. The first is a MI approach with no auxiliary variables, the second is a MI model with auxiliary variables with 10 imputations, the final model is a MI model with auxiliary variables and 100 imputations. For the auxiliary variables 15 were used, 10 were constructed only related to the missingness and another 5 were constructed that were related to the missingness and the variable itself. All three MI models were identical to the ‘god’ model. This demonstrates that the simulation is acting as it should and also demonstrates an MCAR mechanism. Overall if there is confidence that the data is MCAR handling missing data practices should return the same or near identical substantive results as your complete records analysis.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1: Simulation Regression Models Using a MCAR 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** | -0.18 | \*\*\* | -0.18 | \*\*\* | -0.18 | \*\*\* | -0.26 | \*\*\* | -0.26 | \*\*\* | -0.18 | \*\*\* | -0.18 | \*\*\* | -0.17 | \*\*\* | -0.18 | \*\*\* | -0.18 | \*\*\* |
|  | (0.02) |  | (0.02) |  | (0.02) |  | (0.01) |  | (0.01) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  |
| **Independent Variable 2** | -0.19 | \*\*\* | -0.19 | \*\*\* | -0.20 | \*\*\* | -0.26 | \*\*\* | -0.26 | \*\*\* | -0.20 | \*\*\* | -0.19 | \*\*\* | -0.19 | \*\*\* | -0.20 | \*\*\* | -0.20 | \*\*\* |
|  | (0.02) |  | (0.02) |  | (0.02) |  | (0.01) |  | (0.01) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  |
| **Independent Variable 3** | -0.19 | \*\*\* | -0.19 | \*\*\* | -0.20 | \*\*\* | -0.06 | \*\*\* | -0.06 | \*\*\* | -0.20 | \*\*\* | -0.20 | \*\*\* | -0.20 | \*\*\* | -0.19 | \*\*\* | -0.19 | \*\*\* |
|  | (0.02) |  | (0.02) |  | (0.02) |  | (0.01) |  | (0.01) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  |
| **Intercept** | 1.15 | \*\*\* | 1.15 | \*\*\* | 1.16 | \*\*\* | 1.29 | \*\*\* | 1.31 | \*\*\* | 1.16 | \*\*\* | 1.15 | \*\*\* | 1.15 | \*\*\* | 1.16 | \*\*\* | 1.16 | \*\*\* |
|  | (0.02) |  | (0.02) |  | (0.03) |  | (0.02) |  | (0.01) |  | (0.03) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  |
| **Number of observations** | 1000 |  | 1000 |  | 512 |  | 1000 |  | 1000 |  | 512 |  | 1000 |  | 1000 |  | 1000 |  | 1000 |  |
| **AIC** | -1245.53 |  | 819.95 |  | -649.50 |  | -1125.29 |  | -1125.41 |  | -649.50 |  | 825.34 |  |  |  |  |  |  |  |
| **BIC** | -1225.90 |  | 844.49 |  | -632.55 |  | -1105.65 |  | -1105.78 |  | -632.55 |  | 894.05 |  |  |  |  |  |  |  |
| **Adjusted R-squared** | 0.80 |  |  |  | 0.81 |  | 0.78 |  | 0.78 |  | 0.81 |  |  |  | . |  | . |  | . |  |
| **\*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: Simulation using a MCAR principle. 51 per cent missingness introduced.** | | | | | | | | | | | | | | | | | | | | |

Duplicating this simulation using the same seed and variable construction but using a MAR mechanism of missingness allows a true assessment of handling missing data methods. Results are presented in table XXXX. As with the MCAR models the ‘God’ model and the Complete SEM model identical results are obtained. Introducing MAR missingness at independent variable three demonstrates that the substantive interpretation of the model does change as would be expected with a MAR mechanism. Moving to the ‘bad’ ways to handle missing data with all missingness coded as being equal to zero and the substantive results overcorrect for missingness. For the second model coding all missingness equal to one, the model has completely different substantive results compared to the ‘God’ model – for the interpretation of independent variable three the interpretation is no longer significant and changes from negative to positive. The Single use modal imputation presents identical results to the prior model due to the dichotomous nature of the independent variable – it too demonstrates its lack of handling missing data. Moving on to the ‘good’ handling practices, the FIML model recovers much of the missingness and whilst not exactly identical in terms of coefficients the substantive interpretation is identical. The same can be said across the three MI models.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2: 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** | -0.18 | \*\*\* | -0.18 | \*\*\* | -0.11 | \*\*\* | -0.16 | \*\*\* | -0.27 | \*\*\* | -0.27 | \*\*\* | -0.21 | \*\*\* | -0.17 | \*\*\* | -0.17 | \*\*\* | -0.17 | \*\*\* |
|  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.01) |  | (0.01) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  |
| **Independent Variable 2** | -0.19 | \*\*\* | -0.19 | \*\*\* | -0.12 | \*\*\* | -0.17 | \*\*\* | -0.27 | \*\*\* | -0.27 | \*\*\* | -0.22 | \*\*\* | -0.18 | \*\*\* | -0.19 | \*\*\* | -0.19 | \*\*\* |
|  | (0.02) |  | (0.02) |  | (0.02) |  | (0.01) |  | (0.01) |  | (0.01) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  |
| **Independent Variable 3** | -0.19 | \*\*\* | -0.19 | \*\*\* | -0.14 | \*\*\* | -0.23 | \*\*\* | 0.03 |  | 0.03 |  | -0.16 | \*\*\* | -0.21 | \*\*\* | -0.20 | \*\*\* | -0.20 | \*\*\* |
|  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.02) |  |
| **Intercept** | 1.15 | \*\*\* | 1.15 | \*\*\* | 0.81 | \*\*\* | 1.11 | \*\*\* | 1.29 | \*\*\* | 1.29 | \*\*\* | 1.23 | \*\*\* | 1.12 | \*\*\* | 1.13 | \*\*\* | 1.14 | \*\*\* |
|  | (0.02) |  | (0.02) |  | (0.04) |  | (0.02) |  | (0.02) |  | (0.02) |  | (0.03) |  | (0.02) |  | (0.02) |  | (0.02) |  |
| **Number of observations** | 1000 |  | 1000 |  | 513 |  | 1000 |  | 1000 |  | 1000 |  | 1000 |  | 1000 |  | 1000 |  | 1000 |  |
| **AIC** | -1245.53 |  | 819.95 |  | -696.98 |  | -1290.15 |  | -1098.29 |  | -1098.29 |  | 792.03 |  |  |  |  |  |  |  |
| **BIC** | -1225.90 |  | 844.49 |  | -680.02 |  | -1270.52 |  | -1078.66 |  | -1078.66 |  | 860.74 |  |  |  |  |  |  |  |
| **Adjusted R-squared** | 0.80 |  |  |  | 0.31 |  | 0.81 |  | 0.77 |  | 0.77 |  |  |  | . |  | . |  | . |  |
| **\*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: Simulation using a MAR principle. 51 per cent missingness introduced.** | | | | | | | | | | | | | | | | | | | | |

So far it has been demonstrated that there are two distinct groups of handling missing data practices. The ‘bad’ ways related to single mean/modal imputation and coding missing data back into the model at arbitrarily assigned values. These handling missing data methods are little more than guess work that rarely recapture the same substantive interpretation as the ‘God’ model. The ‘Good’ ways of handling missing data on the other hand do show a consistent ability to recover the missingness under both a MCAR and MAR mechanism. In the results presented here it appears at first that the MI model with zero auxiliary variables actually does slightly better than a FIML model. This exact simulation was repeated 10 times using 10 different seeds to assess both if this is an accurate conclusion to reach and to confirm the prior findings mentioned above. Whilst the repeated seeds do confirm that FIML and MI approaches are the only handling missing data methods used that accurately present substantive interpretations that are near identical to the ‘god’ model, it appears that the FIML model on average outcompetes a MI model with zero auxiliary variables but there is substantively no difference between using FIML or MI based approaches so long as the MI model has appropriate auxiliary variables.

This simulation was useful in understanding and comparing handling missing data methods. Considering there are no real differences between a FIML or a MI approach, either approach is sensible though for the entirety of this thesis MI approaches will be used over FIML. FIML approaches using a categorical dependent variable cannot be easily conducted using Stata. The ‘sem’ command framework does not allow categorical dependent variables (hence why the simulation was conducted using a metric dependent variable). There are two possible solutions to this; either conduct multiple logistic regressions by breaking up the categorial dependent variable and using those multiple models, or move to a different software that can compute a FIML model with a categorical dependent variable. Only MPlus offers software capable of doing such a task. Neither of these solutions are particularly attractive. The first means that whilst analysis up until this point has been conducted using one singular multinominal logistic regression, now there would instead have to be multiple logistic regressions compared to a complete records analysis that is a multinominal logistic regression. The second solution would require a break in the overall workflow that increases the risk of mistake/bias into the analysis. For these practical, rather that methodological reasons, MI based approaches are adopted to handle missing data in this thesis going forward.

Number of Imputations in MI Models

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.

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[[15]](#footnote-15). 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 1.22. 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, 8,448 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 2.22 Missing data patterns for NCDS

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

Of the missing data, 63 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. The gold standard for dealing with MAR-related data is MI (Treiman, 2009). 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.

### Multiple Imputation by Chained Equations

Multiple Imputation by Chained Equations is a tool developed to address missing data on all variables within a given model simultaneously. It does this by filling in missing values in multiple variables iteratively by using chained equations. Multiple imputation models are estimated using the mi suite in Stata. This suite is compatible with the svy suite and can be adjusted for complex survey design.

While multiple imputation does help with missingness, it has some drawbacks. Goodness-of-fit statistics, for example, cannot be used – and BIC are the most prominent. Therefore, it is not possible to assess the more appropriate or parsimonious model – it is simply possible to compare the substantive effects between a complete records analysis and a multiple imputation model. 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).

Choosing the number of imputations is difficult. Previous literature suggests that anywhere between 3-5 imputations is sufficient to obtain acceptable properties (Carpenter and Kenward, 2012). However, some modern literature suggests closer to 50 imputations (Silverwood *et al.*, 2021). However, if there is a desire to estimate small p-values or have an MI estimator of the fraction of missing information, greater imputations are required. Carpenter and Kenward (2012) suggest two routes. If an analysis after imputation is clear-cut after a small number of imputations, there is no need to perform more. If, however, after imputation, the inference is less clear-cut, take K = 100, or 100 imputations. Others promote a slightly different interpretation. White et al. (2010) and Bodner (2008) suggest using the Fraction of Missing Information (FMI) as a baseline for the minimum required imputations. If the maximum FMI in a given 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 considerably reduced and stabilised.

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.

Auxiliary variables are variables in the data set that are either correlated with a missing variable or variables but are not a part of the primary analytical model of interest. They are included within the imputation model to increase accuracy and statistical power to make the MAR assumption more plausible. Making the MAR assumption more plausible is done by including auxiliary variables – variables that can predict missingness on a given variable. 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).

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 chapter. The imputation model also includes several auxiliary variables to add to maximising the plausibility of the MAR assumption in order 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[[16]](#footnote-16).

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[[17]](#footnote-17) were checked to see the distributions of each variable against the imputations. These graphs are seen below from Figure 1.10-13. Note that due to the sex variable having zero missingness, no graph was produced, as no imputations on that variable were required. As illustrated, all analytical variables that were imputed have a relatively stable mean and standard deviation across the iteration numbers.

A graph showing different colored lines

Description automatically generated with medium confidence

Figure 1.10 Trace plot summaries for Economic Activity

A graph showing different colored lines

Description automatically generated

Figure 1.11 Trace plot summaries for Educational Attainment

A graph showing different colored lines

Description automatically generated with medium confidence

Figure 1.12 Trace plot summaries for NS-SEC

A graph showing the number of numbers

Description automatically generated with medium confidence

Figure 1.13 Trace plot summaries for Housing Tenure

The following models presented will compare a complete records analysis using NS-SEC from the previous chapter and the imputed model in Table 1.15. The CRA model has 7,915 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.

The results for both the complete records analysis and the imputed model can be viewed in Table 1.23. 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 largest single difference in average marginal effects between the CRA and the imputed model amounts to 1 per cent – a difference that ultimately does not change the substantive interpretation of the overall 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. At this point 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.

Table 2.23 Comparison of CRA NS-SEC vs Imputed NS-SEC

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CRA NS-SEC | | | Average Marginal Effects | | Imputed NS-SEC | | | Average Marginal Effects | |
| Economic Activity | **Coef.** | **S.E** | **Sig.** | **Prob.** | **S.E** | **Coef.** | **S.E** | **Sig.** | **Prob.** | **S.E** |
| Employment |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.58 | (0.08) | \*\*\* | -0.39 | (0.01) | -3.52 | (0.07) | \*\*\* | -0.39 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | -0.40 | (0.07) | \*\*\* | -0.17 | (0.01) | -0.44 | (0.06) | \*\*\* | -0.17 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.68 | (0.08) | \*\*\* | 0.08 | (0.01) | 0.69 | (0.07) | \*\*\* | 0.09 | (0.01) |
| NS-SEC |  |  |  |  |  |  |  |  |  |  |
| *1.1* | 0.08 | (0.23) |  | 0.02 | (0.03) | -0.07 | (0.22) |  | 0.00 | (0.04) |
| *1.2* | -0.67 | (0.22) | \*\* | -0.05 | (0.03) | -0.73 | (0.21) | \*\*\* | -0.06 | (0.03) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3* | 0.36 | (0.15) | \* | 0.05 | (0.02) | 0.35 | (0.15) | \* | 0.05 | (0.02) |
| *4* | 0.96 | (0.14) | \*\*\* | 0.08 | (0.02) | 0.90 | (0.14) | \*\*\* | 0.08 | (0.02) |
| *5* | 0.87 | (0.14) | \*\*\* | 0.07 | (0.02) | 0.84 | (0.14) | \*\*\* | 0.08 | (0.02) |
| *6* | 0.91 | (0.14) | \*\*\* | 0.09 | (0.02) | 0.88 | (0.14) | \*\*\* | 0.09 | (0.02) |
| *7* | 1.36 | (0.13) | \*\*\* | 0.13 | (0.02) | 1.39 | (0.14) | \*\*\* | 0.14 | (0.02) |
| Intercept | 0.89 | (0.12) | \*\*\* | (.) | (.) | 0.86 | (0.11) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Non-Traditional Education |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -1.71 | (0.09) | \*\*\* | 0.03 | (0.01) | -1.68 | (0.08) | \*\*\* | 0.02 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | -0.84 | (0.09) | \*\*\* | -0.07 | (0.01) | -0.86 | (0.08) | \*\*\* | -0.07 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | -0.11 | (0.10) |  | -0.04 | (0.01) | -0.17 | (0.09) |  | -0.04 | (0.01) |
| NS-SEC |  |  |  |  |  |  |  |  |  |  |
| *1.1* | 0.02 | (0.23) |  | 0.00 | (0.02) | -0.07 | (0.23) |  | -0.00 | (0.02) |
| *1.2* | -0.52 | (0.21) | \* | -0.01 | (0.02) | -0.47 | (0.19) | \* | -0.01 | (0.02) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3* | 0.08 | (0.16) |  | -0.01 | (0.01) | 0.06 | (0.16) |  | -0.01 | (0.01) |
| *4* | 0.15 | (0.17) |  | -0.03 | (0.01) | 0.15 | (0.16) |  | -0.03 | (0.01) |
| *5* | 0.24 | (0.15) |  | -0.02 | (0.01) | 0.22 | (0.15) |  | -0.02 | (0.01) |
| *6* | -0.02 | (0.16) |  | -0.04 | (0.01) | 0.00 | (0.16) |  | -0.04 | (0.01) |
| *7* | 0.44 | (0.15) | \*\* | -0.03 | (0.01) | 0.48 | (0.15) | \*\* | -0.03 | (0.01) |
| Intercept | 0.18 | (0.13) |  | (.) | (.) | 0.13 | (0.12) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| School | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Training & Apprenticeships |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.24 | (0.09) | \*\*\* | -0.17 | (0.01) | -3.20 | (0.08) | \*\*\* | -0.16 | 0.01 |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 1.52 | (0.09) | \*\*\* | 0.24 | (0.01) | 1.52 | (0.08) | \*\*\* | 0.24 | 0.01 |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.38 | (0.08) | \*\*\* | -0.01 | (0.01) | 0.35 | (0.08) | \*\*\* | -0.02 | 0.01 |
| NS-SEC |  |  |  |  |  |  |  |  |  |  |
| *1.1* | -0.12 | (0.26) |  | -0.02 | (0.03) | -0.13 | (0.25) |  | -0.01 | (0.03) |
| *1.2* | -0.45 | (0.23) |  | 0.00 | (0.03) | -0.50 | (0.23) | \* | -0.00 | (0.02) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3* | 0.19 | (0.17) |  | -0.00 | (0.02) | 0.13 | (0.17) |  | -0.01 | (0.02) |
| *4* | 0.97 | (0.16) | \*\*\* | 0.05 | (0.02) | 0.91 | (0.17) | \*\*\* | 0.04 | (0.02) |
| *5* | 0.90 | (0.15) | \*\*\* | 0.05 | (0.02) | 0.86 | (0.15) | \*\*\* | 0.04 | (0.02) |
| *6* | 0.84 | (0.15) | \*\*\* | 0.04 | (0.02) | 0.81 | (0.15) | \*\*\* | 0.03 | (0.02) |
| *7* | 1.08 | (0.15) | \*\*\* | 0.02 | (0.02) | 1.07 | (0.15) | \*\*\* | 0.02 | (0.02) |
| Intercept | -0.72 | (0.14) | \*\*\* | (.) | (.) | -0.80 | (0.13) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Unemployment & Out of Labour Force |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.94 | (0.27) | \*\*\* | -0.03 | (0.00) | -4.04 | (0.22) | \*\*\* | -0.04 | (0.00) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | -0.37 | (0.14) | \*\* | -0.01 | (0.00) | -0.29 | (0.11) | \*\* | -0.01 | (0.00) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.8 | (0.16) | \*\*\* | 0.01 | (0.00) | 0.94 | (0.14) | \*\*\* | 0.02 | (0.00) |
| NS-SEC |  |  |  |  |  |  |  |  |  |  |
| *1.1* | -0.35 | (0.67) |  | -0.01 | (0.02) | -0.33 | (0.57) |  | -0.01 | (0.02) |
| *1.2* | -2.08 | (1.03) | \* | -0.03 | (0.01) | -2.14 | (1.05) | \* | -0.03 | (0.01) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3* | -0.29 | (0.38) |  | -0.01 | (0.01) | -0.09 | (0.36) |  | -0.01 | (0.01) |
| *4* | 0.60 | (0.31) |  | -0.00 | (0.01) | 0.61 | (0.29) | \* | -0.00 | (0.01) |
| *5* | 0.29 | (0.30) |  | -0.01 | (0.01) | 0.24 | (0.28) |  | -0.01 | (0.01) |
| *6* | 0.77 | (0.28) | \*\* | 0.00 | (0.01) | 0.73 | (0.28) | \* | 0.00 | (0.01) |
| *7* | 1.20 | (0.27) | \*\*\* | 0.01 | (0.01) | 1.27 | (0.27) | \*\*\* | 0.01 | (0.01) |
| Intercept | -1.48 | (0.25) | \*\*\* | (.) | (.) | -1.47 | (0.24) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Number of observations | 8411 | | | | | 12536 | | | | |
| Average RVI |  | | | | | 0.31 | | | | |
| Largest FMI |  | | | | | 0.40 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: Comparison of CRA NS-SEC vs Imputed NS-SEC model | | | | | | | | | | |

Additional checks on the imputed model are produced via post-estimation statistics through RVI and FMI. A high RVI corresponds to large amounts of missing data and/or that they are weakly correlated with other variables within the imputed model. The closer the RVI is to zero, the less effect missing data has on a variable’s variance. The FMI, on the other hand, relates to the proportion of the total sampling variance due to missing data. The higher the FMI is, the greater the number of imputations required for reliable results. The number of imputations should be roughly equivalent to the highest FMI percentage (which has been followed in this model). The highest FMI in the imputed model is 44 per cent, indicating a requirement of at least 44 imputations – the MI model has 50 imputations. Relative efficiency indicates efficiency. The closer it is towards one indicates that the analysis has the correct number of imputations.

The average RVI score was 0.31, meaning that, on average, the missing data has a negligible impact on the model’s variance. According to the RVI scores across categories, NS-SEC across all economic activity and tenure within unemployed and out of the labour force are the only two analytical variables that have consistently above average (greater than 0.30) RVI scores. For all other analytical variables, missing data has little effect on their variance. Housing tenure and NS-SEC have moderate (greater than 0.25) FMI scores, indicating that a substantive amount of the total sampling variance is due to missing data. The FMI value of Intermediate occupations for unemployed & OLF is the highest FMI value from the model with a value of 0.36. This value indicates that 50 imputations was an appropriate number of imputations. All variables have a relative efficiency close to one – none have a relative efficiency below 0.99. This lends support to the notion that 50 imputations are an appropriate number.

Regarding parameter-specific degrees of freedom (DF) and percentages of increase in standard errors due to nonresponse. The closer RVI is to zero, the larger the degrees of freedom, regardless of the number of imputations. The highest degrees of freedom correspond to sex. This suggests that the loss of information due to non-response is the smallest for this analytical variable. This is unsurprising, given the level of missingness related to sex.

### Discussion and Conclusions

The fact 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. In essence, 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. Going forward, 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 Part 1

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 transitions from mandatory education. 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.

More socio-economically advantaged youth tend to maintain a pathway of elongated education, compared to their less advantaged peers, who are more likely to enter employment. There is a caveat within this however, pathways expressed in non-traditional educational pathways as well as training & apprenticeship programs offer an opportunity for those at the lower end of socio-economic advantage. Most prominently with the latter of these pathways, individuals whose fathers are from skilled manual backgrounds are most likely to take up this opportunity – though these individuals are most likely to be men rather than women. This finding is consistent with previous studies related to the class divide within training & apprenticeships (Booth and Satchell, 1994) and the sex divide within the segregated labour market of the time (Dex and Bukodi, 2012).

Sex based differences within pathways are most evident within this training & apprenticeship category, although differences are also apparent with women more likely to enter employment over men. Whilst traditionally masculine fields dominated training & apprenticeship programs during the timeframe of the NCDS, the explanation as to why more women than men entered straight into employment is slightly more nebulous. One possibility relates to rational choice theory (Goldthorpe, 1998); given their structural position, women saw entering employment directly after mandatory schooling as the most secure and stable pathway for their given life course.

Other structural factors like housing tenure also matter – but not to the extent of traditional structural inequalities such as social class and sex. These findings of structural effects impacting life chances are found in previous literature (Saunders, 2021). This empirical evidence suggests that over and above matters of social class, how you live – in the case of housing tenure – presents a transparent barrier to entry regarding the choices and opportunities individuals make 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. One aspect of these structural effects that have not been reflected within this research is the recency onto which things like housing tenure and social class position may have upon constricting choice and opportunity. There is a possibility that an individual in a household that rented for most of their life and only recently bought their home would have a different magnitude of effect compared to an individual whose household had always owned their own home. Similarly, an individual who has had multi-generational social class stability may have a different magnitude of effect compared to an individual whose family has very recently experienced upward (or downward) social mobility. This does not fit the purview of this research but is certainly something not to be considered for future research within this area.

The findings from this empirical research appear to confirm the relatively 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. Most notably, these affect whether an individual seeks to continue along a traditional educational pathway. In saying this, lack of educational success at 16 does not block youth from entering several pathways. It appears to influence traditional education, but this is not true for employment, non-traditional educational pathways, and training & apprenticeships. This demonstrates that less academically able youth desire alternative pathways beyond employment and that, given the socio-historical context of the NCDS cohort, the option was there to allow young people to enter these alternative pathways.

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 and a slight preference concerning AIC statistics. Through its implementation, social class was found to have a resounding impact on individuals’ choices and opportunities concerning transitional experiences. 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.

The results are also innovative by assessing missingness within the complete records analysis model. 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.

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 replicate 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.

# The British Cohort Study: Youth Transitions in Education and Employment

"a different country… You have to blink and rub your eyes". – Jacques (1982)[[18]](#footnote-18)

## Introduction

Continuing the theme set in Chapter One, Chapter Two attempts to replicate the previous analysis of entry from school into work. This chapter focuses on the British Cohort Study (BCS) that started in 1970. As in chapter one, this chapter will focus on the pathways and choices made by individuals in the BCS after they reached 16 and ended mandatory schooling. Once more, focus will be placed on structural inequalities of social class, sex, and housing tenure to understand young people’s transitional experiences during this period. An attempt is made to duplicate the NCDS chapter as precisely as possible to enable a comparison of cohort transitional experiences.

This chapter begins with a literature review of the present literature on BCS youth and their transitional experiences. Then, this chapter will move on to a duplication analysis of the model used in chapter one, with sensitivity analysis and handling missing data sections. Finally, this chapter will conclude with a comparison of NCDS and BCS youth.

The data was used from sweeps from birth to age 30 using the BCS. The present chapter continues the tradition of the previous by updating prior literature with modern statistical techniques. This chapter will look at four primary economic activity outcomes: employment, education, training & apprenticeship, and unemployment & out of the labour force. Due to the nature of the BCS, missingness will form a critical discussion within this chapter.

## Literature Review: BCS Timeframe and Context

This section provides an overview of the literature within the field of youth transitions of the BCS cohort. This review focuses on existing research outlining the school-to-work transition and examining the structural impacts of that transition within the context of the BCS cohort. Initially, the literature will focus on the historical and temporal context of the BCS cohort to ground the empirical research on transitions. As with Chapter One, major transition themes are identified as they relate to employment, education, training, and unemployment. Each is influenced to some degree by structural factors that impact individual choice and opportunity. The changing nature of the labour market and British polity during the BCS timeframe 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.

At the time of the BCS cohort, 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.

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 made 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).

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 (MSC). This in turn would be replaced by Youth Training (YT) in 1990 (Droy, Goodwin and O’connor, 2019). 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 opernadi 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). 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 (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. 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 that did not enter the YTS. For women, the effects on earnings were small and insignificant (Dolton, Galinda-Rueda and Makepeace, 2004).

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).

All this historical phenomenon has impacted the relative stability of youth transitions, which is apparent for the NCDS cohort. The relative decline in individuals moving straight from school into work after mandatory schooling and the growth of tricky transitions and accumulating human capital via higher education suggests increased risk and uncertainty (Anders and Dorsett, 2017).

### Structural Barriers to successful transitions – the role of social class and sex

#### 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).

#### 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).

#### 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.

## Data and Methods

Chapter Two is a replication analysis of the models presented in Chapter One using the NCDS. Therefore, similar to Chapter One, the relationship between social origins and economic activity after mandatory schooling is examined using the large-scale, nationally representative data collected from the British Cohort Study. Educational attainment, housing tenure, and sex are also included in the model, as they were in the NCDS model. This is to assess choice and opportunity into different forms of economic activity: employment, education, training & apprenticeships, and unemployment & out of the labour force. BCS data is available using the UK Data Service.

Before any modelling, it is essential to note that 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.

Whilst this chapter has attempted to replicate the analysis in chapter one, there are some substantive differences. Firstly, and most substantially, chapter two's outcome variable of economic activity after mandatory schooling only has four categories in chapter five. Chapter two is missing a ‘post-education schooling’ category that encapsulates non-traditional forms of education that did not follow the traditional university route. For the BCS cohort, these non-traditional forms of education had decreased in popularity despite not being appropriately recorded In the BCS survey. The second substantive change relates to the construction of social class measures (NS-SEC and RGSC) within chapter two. Whilst both chapter one and chapter two use occupational coding data from 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.

As with Chapter One, after an initial exploration of descriptive statistics, multinominal logistic regression will be used to understand the choices and opportunities of BCS youth regarding economic activity post-mandatory schooling. After this initial model, a sensitivity analysis of social stratification measures will be employed to assess the most appropriate measure. Finally, an analysis of missing data involving multiple imputation will be conducted to evaluate the impact of missingness on the substantive findings of the model.

Sample Attrition and missingness:

Table 3.1 Participation in the BCS from Birth to 30 years

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Total cohort | Dead | Permanent Emigrants | Non-Response | Other[[19]](#footnote-19) | Participants | (% of the eligible sample)[[20]](#footnote-20) | 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[[21]](#footnote-21) | 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[[22]](#footnote-22) | 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.

### Introduction to measures for subsequent analysis

The following section provides an overview of key variables used for subsequent multivariate analysis. Variable selection was a process that involved a combination of using the CLOSER search platform, digital codebooks, and manual searching of the individual BCS databases. This search was much more manageable in Chapter Two than in Chapter One, as the analytical variables in question were already chosen in Chapter One. Thus, the goal of variable selection in Chapter Two was to find the most appropriate similar measurements used in the BCS.

#### Economic Activity

As the primary purpose of this chapter is to replicate the analysis of Chapter One with a different cohort from an additional period, the economic activity outcome variable has been selected in much the same way it was for Chapter One. The variable of interest was economic activity data for individuals of month 201 –when participants were 16 in September. As the BCS cohort’s mandatory schooling period ended when participants were 16, this was a natural month selection to measure economic activity. The month of September gives enough time for O-level results to be received and have any potential impact on an individual’s economic activity circumstances.

Economic activity was recorded retrospectively in the 21-year-old sub-sweep. It 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 XXXX. Some recoding was required. The economic activity variable has four outcomes: employment, education, training & apprenticeships, and unemployment & out of the labour force. Employment is defined as any individual who, after mandatory education, entered employment. Education was coded as any individual that stayed within some form of schooling or education post-mandatory period; this for most would be individuals taking A-level examinations. The training & apprenticeships category is coded as all those training schemes that do not fall under education. Finally, the unemployment and out of the labour force category codes anyone who is either unemployed, out of the labour force, or otherwise economically inactive. The response level for this last category is minimal and can potentially impact the statistical power of the model adversely. Unfortunately, there is no viable alternative to this.

Table 3.2 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 |

This economic activity variable was re-coded along similar lines within the NCDS model. Unlike the NCDS model, there is no qualitative distinction between what specific type of education an individual engages in as a form of economic activity. There is only one ‘FT Education’ category. Therefore, where the NCDS model had five distinct outcomes within the economic activity outcome variable, the BCS model will only have four: Employment, Education, Training & Apprenticeship, and Unemployment & Out of the Labour Force. The categories ‘FT Employee’, ‘PT Employee’, and ‘Self-employed’ were combined to make an ‘Employment’ category. The category ‘FT Education’ was renamed the ‘Education’ category. The category ‘On Training Scheme’ was renamed the ‘Training & Apprenticeship’ category. Finally, the categories ‘Seeking Work’ and ‘Looking After Children/home ft’ were combined to make an ‘Unemployed & Out of the Labour Force’ category. The ‘Something Else’ category from the raw economic activity variable was re-coded as missing because it was too vague to be placed in other outcome categories. The recorded economic activity variable [econ201] thus has a total frequency of 1,618.

#### 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). By being the previous cohort to share this, the variable dictated as educational attainment is directly replicated from chapter one. 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. At that point in the cohort, only 9,003 participants responded; of those, only 5,438 responded to an educational attainment variable. The BCS documents O’level attainment in two ways. The first is a variable of the number of O’level passes ranging from A-C grade [b960169]. 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 former variable 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.

Compared to the NCDS cohort, most BCS cohorts gained five or more O-level passes. The table below illustrates that more of the BCS cohort gained five or more O-level passes proportionally compared to the NCDS cohort.

[insert comparison of O-level passes for NCDS and BCS]

#### Sex

Sex is a variable taken from the birth sweep to ensure the most significant number of responses. Sex measures the respondent's sex in a binary male/female format, as seen in chapter one; sex as a variable played a critical analytical role in understanding the structural inequalities and barriers that play an essential role in choice and opportunity for youth. Its inclusion in the BCS model is essential not just for replication, but as seen in the literature review, the role of women and men in the labour market was still undergoing systematic changes. 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].

#### Housing Tenure

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’ (XXXX) 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] – similar to 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 similar to 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.

#### Social Stratification and Socio-Economic Background: NS-SEC, CAMSIS, RGSC

As in Chapter One, a core component of this chapter is conducting a sensitivity analysis of social stratification measures. The continuation of a sensitivity analysis in Chapter Two provides a basis for comparison with Chapter One. As seen in Chapter One, each of the three models that used a different social stratification measure was relatively and substantively identical. Chapter two seeks to understand whether that is a unique phenomenon for the NCDS cohort or a pattern replicating in different periods.

As mentioned, Chapter Two's RGSC and NS-SEC measures differ slightly from their Chapter One counterparts. 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 of these 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 [B3FSSOC90] and the mother’s data [B3MSSOC90]. 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 SOC90 codes provided by Gregg (ibid).

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 (XXXX). Below is a table comparing each social stratification measure between the NCDS and BCS cohorts.

##### SOC Codes

As with the NCDS, the BCS dataset offers occupational SOC codes at the 2000 and 90 level. These SOC codes will be used to construct each instance of social stratification: NS-SEC, CAMSIS, and RGSC in order to assess the similarities and differences between and within social stratification measures. Another goal of this sensitivity analysis using SOC codes will be to assess the degree of difference across databases – seeking to understand if a database such as the BCS that started in 1970 will have an overall similar SOC 90 and SOC 2000 models compared to a database such as the NCDS that started in 1958.

The following descriptive statistics will include three measures of social stratification: NS-SEC, CAMSIS, and RGSC followed by their SOC 2000 and SOC 90 constructions. Differences will first be recorded on the descriptive level, followed by a comparison of analytical models after the sensitivity analysis of social stratification measures is conducted.

## Descriptive Statistics

Table 2.4 shows the frequencies and summary statistics for the BCS. Overall, 16.78 per cent of the sample is in employment. Whilst 64.04 per cent are in education – making up the majority of the sample. Regarding training & apprenticeships, 16.32 per cent of respondents are in this form of economic activity. Only 2.77 per cent of respondents are unemployed & out of the labour force – this is potentially the influence of the YTS.

Regarding educational attainment, 57.26 per cent of individuals received less than five O-levels, while the remaining 42.74 per cent received five or more O-levels. Sex illustrates a slight overrepresentation of men (57.68 per cent) compared to women (42.32 per cent). Regarding home ownership, 21.02 per cent of individuals grew up in a home that wasn’t owned by their parents compared to 78.98 per cent that did.

The NS-SEC SOC 2000 categories all see a relatively even distribution of respondents between 10-20 per cent except for NS-SEC 1.1 and 1.2. The relatively low number of observations in both categories has the possibility to cause issues related to statistical power within the model. Due to these categories being split over four outcome categories, the standard errors for these categories will most likely be very high. The level of interpretation that can be gained from these specific categories within NS-SEC will be low. The NS-SEC SOC 90 construction follows a very similar distribution to that of its SOC 2000 counterpart.

RGSC SOC 2000 is much more unevenly distributed in comparison to its NS-SEC counterpart. Skilled manual occupations comprise 34.44 per cent of respondents, with professional and unskilled occupations making up 5.53 and 4.43 per cent, respectively. Similar to comments about NS-SECs' intermediate occupations, the same can be said about RGSCs' professional occupations. The RGSC SOC 90 construction follows a similar pattern to its SOC 90 counterpart with a slight increase in Professional occupations – going from 5.53 per cent in SOC 2000 to 7.33 per cent in SOC 90. Overall, except some small adjustments, both the SOC 2000 and SOC 90 distributions follow a similar trend.

CAMSIS, as a metric measure, does not have issues related to statistical power that have been mentioned in relation to NS-SEC. The CAMSIS SOC 2000 construction has a mean of 49.71 and a standard deviation of 13.82. The CAMSIS SOC 90 construction has a mean of 50.13 and a standard deviation of 14.53. Overall, these two measures are remarkably similar and so the substantive results from a comparison of two models using these measures should present results near identical to one another.

Table 3.3 Descriptive Statistics for Economic Activity Model

|  |  |  |
| --- | --- | --- |
| Table 1: Descriptive Statistics for Economic Activity | | |
|  | n | % |
| Economic Activity of Respondent |  |  |
| *Employment* | 122 | 16.87% |
| *Education* | 463 | 64.04% |
| *Training & Apprenticeships* | 118 | 16.32% |
| *Unemployment & OLF* | 20 | 2.77% |
| Educational Attainment O'levels |  |  |
| *Less than Five O'Levels* | 414 | 57.26% |
| *Five or More O'Levels* | 309 | 42.74% |
| Sex of Respondent |  |  |
| *Female* | 417 | 57.68% |
| *Male* | 306 | 42.32% |
| Housing Tenure of Respondent when a Child |  |  |
| *Own Home* | 571 | 78.98% |
| *Don't Own Home* | 152 | 21.02% |
| Semi-Dominant NS-SEC Social Class of Parents when Respondent was 10 SOC2000 |  |  |
| *Large Employers and higher managerial occupations* | 42 | 5.81% |
| *Higher professional occupations* | 55 | 7.61% |
| *Lower Managerial and professional occupations* | 142 | 19.64% |
| *Intermediate occupations* | 82 | 11.34% |
| *Small employers and own account workers* | 72 | 9.96% |
| *Lower supervisory and technical occupations* | 125 | 17.29% |
| *Semi-routine occupations* | 89 | 12.31% |
| *Routine occupations* | 116 | 16.04% |
| Semi-Dominant NS-SEC Social Class of Parents when Respondent was 10 SOC90 |  |  |
| *Large Employers and higher managerial occupations* | 41 | 5.67% |
| *Higher professional occupations* | 61 | 8.44% |
| *Lower Managerial and professional occupations* | 142 | 19.64% |
| *Intermediate occupations* | 85 | 11.76% |
| *Small employers and own account workers* | 73 | 10.10% |
| *Lower supervisory and technical occupations* | 127 | 17.57% |
| *Semi-routine occupations* | 90 | 12.45% |
| *Routine occupations* | 104 | 14.38% |
| Semi-Dominant RGSC Social Class of Parents when Respondent was 10 SOC2000 |  |  |
| *Professional* | 40 | 5.53% |
| *Managerial and Technical* | 219 | 30.29% |
| *Skilled non-manual* | 89 | 12.31% |
| *Skilled manual* | 249 | 34.44% |
| *Partly skilled* | 94 | 13.00% |
| *Unskilled* | 32 | 4.43% |
| Semi-Dominant RGSC Social Class of Parents when Respondent was 10 SOC90 |  |  |
| Professional | 53 | 7.33% |
| *Managerial and Technical* | 192 | 26.56% |
| *Skilled non-manual* | 117 | 16.18% |
| *Skilled manual* | 234 | 32.37% |
| *Partly skilled* | 104 | 14.38% |
| *Unskilled* | 23 | 3.18% |
|  |  |  |
|  | Mean | SD |
| Semi-Dominant CAMSIS Respondent was 10 SOC2000 | 49.71 | 13.82 |
| Semi-Dominant CAMSIS Respondent was 10 SOC90 | 50.13 | 14.53 |
|  |  |  |
| n |  | 723 |
| Data Source BCS [Birth-Age30] | | |

From Table 2.5, some observations can be made. Individuals’ economic activity grouping stratifies educational attainment. The only economic activity category with a majority of individuals with five or more O’levels is the continuing education category. For employment, training & apprenticeships, and unemployment & OLF, most individuals received less than five O’levels at school. Concerning respondents’ sex by economic activity, the only category with a majority of male respondents is the unemployment & OLF category – with the employment, education, and training & apprenticeship categories having a majority of women. Concerning housing tenure, all categories see a majority of respondents living in homes their parents owned – though the most significant majority of these categories resides within the education category.

Moving on to NS-SEC SOC 2000 construction, the most relevant observation is that NS-SEC 1.1 and 3 have zero observations in the unemployment & OLF categories. This will pose statistical power problems when modelling economic activity – the standard errors will also be very high. The distribution of NS-SEC changes depending on the economic activity category that is looked at. For example, those who enter education as an economic activity see a more significant proportion of respondents from NS-SEC 1.1-3 compared to all other economic activity categories. Conversely, NS-SEC 7 has a more significant proportion of respondents within employment than education. The SOC 90 construction of NS-SEC broadly shares the same distribution with the SOC 2000 version. That being said, there are some notable differences. Whilst the SOC 2000 version identifies that that NS-SEC 1.2 has its highest concentration in the educational category, for SOC 90 it is instead within the unemployment & OLF category. Within NS-SEC 2, this happens once again. The SOC 2000 construction states that the smallest concentration is within the unemployment & OLF category whereas the SOC 90 construction states it is instead within the training & apprenticeship category. This happens throughout the rest of NS-SEC categories. These examples are used to demonstrate that whilst the SOC 2000 and SOC 90 constructions of NS-SEC share a similar overall distribution, the breakdown of each highlights some substantive differences that may impact upon a model using either measure of social stratification.

Moving on to RGSC SOC 2000 construction, the most relevant observation to be made here is that, like NS-SEC, some categories within economic activity have zero observations and will thus impact statistical power going into the sensitivity analysis. This is true for RGSC 3 non-manual within the unemployment & OLF category. Beyond this, like the NCDS analysis, a straightforward manual/non-manual divide becomes apparent when looking at these descriptive statistics; for the unemployment & OLF category, 85 per cent of respondents reside within manual occupations. Comparatively, over 55 per cent of respondents within the education category reside within non-manual occupations. Echoing the comparison made between SOC 2000 and SOC 90 constructions of NS-SEC, whilst the RGSC measures share an overall similar distribution there are examples of this deviating which may have implications for any analytical model using either social stratification measure – and highlights the need for a sensitivity analysis of different SOC constructions of social stratification measures.

For CAMSIS SOC 2000 construction, there is a base total mean of 49.71. The only economic activity category with a mean above this relates to the education category at 52.19. Employment, training & apprenticeships, and unemployment & OLF categories have a mean CAMSIS below 49.71– the lowest being unemployment & OLF with a mean CAMSIS of 42.58. The same pattern occurs when reflecting upon the means of the CAMSIS SOC 90 construction.

Table 3.4 Descriptive Statistics by Economic Activity

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Descriptive Statistics by Economic Activity | | | | | |
|  | Economic Activity of Respondent | | | | |
|  | Employment | Education | Training & Apprenticeships | Unemployment & OLF | Total |
| Educational Attainment O'levels |  |  |  |  |  |
| *Less than Five O'Levels* | 94.00 (77.05%) | 201.00 (43.41%) | 100.00 (84.75%) | 19.00 (95.00%) | 414.00 (57.26%) |
| *Five or More O'Levels* | 28.00 (22.95%) | 262.00 (56.59%) | 18.00 (15.25%) | 1.00 (5.00%) | 309.00 (42.74%) |
| Sex of Respondent |  |  |  |  |  |
| *Female* | 67.00 (54.92%) | 275.00 (59.40%) | 67.00 (56.78%) | 8.00 (40.00%) | 417.00 (57.68%) |
| *Male* | 55.00 (45.08%) | 188.00 (40.60%) | 51.00 (43.22%) | 12.00 (60.00%) | 306.00 (42.32%) |
| Housing Tenure of Respondent when a Child |  |  |  |  |  |
| *Own Home* | 89.00 (72.95%) | 386.00 (83.37%) | 82.00 (69.49%) | 14.00 (70.00%) | 571.00 (78.98%) |
| *Don't Own Home* | 33.00 (27.05%) | 77.00 (16.63%) | 36.00 (30.51%) | 6.00 (30.00%) | 152.00 (21.02%) |
| Semi-Dominant NS-SEC Social Class of Parents when Respondent was 10 SOC2000 |  |  |  |  |  |
| *Large Employers and higher managerial occupations* | 2.00 (1.64%) | 35.00 (7.56%) | 5.00 (4.24%) | 0.00 (0.00%) | 42.00 (5.81%) |
| *Higher professional occupations* | 5.00 (4.10%) | 46.00 (9.94%) | 2.00 (1.69%) | 2.00 (10.00%) | 55.00 (7.61%) |
| *Lower Managerial and professional occupations* | 21.00 (17.21%) | 107.00 (23.11%) | 13.00 (11.02%) | 1.00 (5.00%) | 142.00 (19.64%) |
| *Intermediate occupations* | 10.00 (8.20%) | 60.00 (12.96%) | 12.00 (10.17%) | 0.00 (0.00%) | 82.00 (11.34%) |
| *Small employers and own account workers* | 12.00 (9.84%) | 38.00 (8.21%) | 17.00 (14.41%) | 5.00 (25.00%) | 72.00 (9.96%) |
| *Lower supervisory and technical occupations* | 25.00 (20.49%) | 69.00 (14.90%) | 27.00 (22.88%) | 4.00 (20.00%) | 125.00 (17.29%) |
| *Semi-routine occupations* | 19.00 (15.57%) | 47.00 (10.15%) | 19.00 (16.10%) | 4.00 (20.00%) | 89.00 (12.31%) |
| *Routine occupations* | 28.00 (22.95%) | 61.00 (13.17%) | 23.00 (19.49%) | 4.00 (20.00%) | 116.00 (16.04%) |
| Semi-Dominant NS-SEC Social Class of Parents when Respondent was 10 SOC90 |  |  |  |  |  |
| *Large Employers and higher managerial occupations* | 1.00 (0.82%) | 36.00 (7.78%) | 4.00 (3.39%) | 0.00 (0.00%) | 41.00 (5.67%) |
| *Higher professional occupations* | 9.00 (7.38%) | 46.00 (9.94%) | 4.00 (3.39%) | 2.00 (10.00%) | 61.00 (8.44%) |
| *Lower Managerial and professional occupations* | 19.00 (15.57%) | 109.00 (23.54%) | 11.00 (9.32%) | 3.00 (15.00%) | 142.00 (19.64%) |
| *Intermediate occupations* | 10.00 (8.20%) | 63.00 (13.61%) | 12.00 (10.17%) | 0.00 (0.00%) | 85.00 (11.76%) |
| *Small employers and own account workers* | 14.00 (11.48%) | 37.00 (7.99%) | 18.00 (15.25%) | 4.00 (20.00%) | 73.00 (10.10%) |
| *Lower supervisory and technical occupations* | 26.00 (21.31%) | 69.00 (14.90%) | 27.00 (22.88%) | 5.00 (25.00%) | 127.00 (17.57%) |
| *Semi-routine occupations* | 19.00 (15.57%) | 46.00 (9.94%) | 21.00 (17.80%) | 4.00 (20.00%) | 90.00 (12.45%) |
| *Routine occupations* | 24.00 (19.67%) | 57.00 (12.31%) | 21.00 (17.80%) | 2.00 (10.00%) | 104.00 (14.38%) |
| Semi-Dominant RGSC Social Class of Parents when Respondent was 10 SOC2000 |  |  |  |  |  |
| Professional | 4.00 (3.28%) | 32.00 (6.91%) | 2.00 (1.69%) | 2.00 (10.00%) | 40.00 (5.53%) |
| *Managerial and Technical* | 28.00 (22.95%) | 167.00 (36.07%) | 23.00 (19.49%) | 1.00 (5.00%) | 219.00 (30.29%) |
| *Skilled non-manual* | 10.00 (8.20%) | 63.00 (13.61%) | 16.00 (13.56%) | 0.00 (0.00%) | 89.00 (12.31%) |
| *Skilled manual* | 52.00 (42.62%) | 135.00 (29.16%) | 48.00 (40.68%) | 14.00 (70.00%) | 249.00 (34.44%) |
| *Partly skilled* | 22.00 (18.03%) | 49.00 (10.58%) | 21.00 (17.80%) | 2.00 (10.00%) | 94.00 (13.00%) |
| *Unskilled* | 6.00 (4.92%) | 17.00 (3.67%) | 8.00 (6.78%) | 1.00 (5.00%) | 32.00 (4.43%) |
| Semi-Dominant RGSC Social Class of Parents when Respondent was 10 SOC90 |  |  |  |  |  |
| *Professional* | 5.00 (4.10%) | 42.00 (9.07%) | 4.00 (3.39%) | 2.00 (10.00%) | 53.00 (7.33%) |
| *Managerial and Technical* | 24.00 (19.67%) | 142.00 (30.67%) | 23.00 (19.49%) | 3.00 (15.00%) | 192.00 (26.56%) |
| *Skilled non-manual* | 15.00 (12.30%) | 85.00 (18.36%) | 17.00 (14.41%) | 0.00 (0.00%) | 117.00 (16.18%) |
| *Skilled manual* | 47.00 (38.52%) | 135.00 (29.16%) | 39.00 (33.05%) | 13.00 (65.00%) | 234.00 (32.37%) |
| *Partly skilled* | 25.00 (20.49%) | 48.00 (10.37%) | 29.00 (24.58%) | 2.00 (10.00%) | 104.00 (14.38%) |
| *Unskilled* | 6.00 (4.92%) | 11.00 (2.38%) | 6.00 (5.08%) | 0.00 (0.00%) | 23.00 (3.18%) |
| Semi-Dominant CAMSIS Respondent was 10 SOC2000 | 46.00 (12.02) | 52.19 (14.15) | 45.00 (11.80) | 42.58 (13.40) | 49.71 (13.82) |
| Semi-Dominant CAMSIS Respondent was 10 SOC90 | 46.48 (12.79) | 52.62 (14.91) | 45.18 (12.46) | 43.89 (13.46) | 50.13 (14.53) |
| N | 122.00 (16.87%) | 463.00 (64.04%) | 118.00 (16.32%) | 20.00 (2.77%) | 723.00 (100.00%) |
|  |  |  |  |  |  |

From table 2.6, a closer inspection of NS-SEC constructed using SOC 2000 and SOC 90 can be viewed from a cross-tabulation of both measures. The majority of respondents from the SOC 2000 construction of NS-SEC share the same NS-SEC category as the SOC 90 construction. The largest share of this relates to NS-SEC 5 with 92.91 per cent. The smallest share relates to NS-SEC 1.2 with 78.69 per cent. There are no extreme surprise displayed within this cross-tabulation. Those that are coded at NS-SEC 3-7 using the SOC 2000 construction of NS-SEC have zero responses coded at NS-SEC 1.1-2 for the SOC 90 construction.

Moving on to table 2.7, a cross-tabulation of RGSC measures can be found. As with the NS-SEC measure, a majority of individuals are coded within the same RGSC categorisation across both measures. The manual/non-manual divide is evident across both measures also. Both measures see a majority of individuals sorted into their respective manual or non-manual distinctions – for example those coded as Routine occupations in the SOC 2000 construction only saw 1 per cent of cases be coded as non-manual – Lower managerial and professional occupations.

Table 2.8 reaffirms earlier statements made about the CAMSIS constructions. The SOC 2000 construction has an overall lower mean than its SOC 90 counterpart whilst having a larger standard deviation than the latter measure. Overall, however, they are both remarkably similar in construction.

Table 3.5 Descriptive Statistics Comparing NS-SEC by SOC2000 and SOC90 Codes

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Semi-Dominant NS-SEC Social Class of Parents when Respondent was 10 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 |
| Semi-Dominant NS-SEC Social Class of Parents when Respondent was 10 SOC2000 |  |  |  |  |  |  |  |  |  |
| *Large Employers and higher managerial occupations* | 37.00 (90.24%) | 0.00 (0.00%) | 0.00 (0.00%) | 3.00 (3.53%) | 2.00 (2.74%) | 0.00 (0.00%) | 0.00 (0.00%) | 0.00 (0.00%) | 42.00 (5.81%) |
| *Higher professional occupations* | 1.00 (2.44%) | 48.00 (78.69%) | 5.00 (3.52%) | 1.00 (1.18%) | 0.00 (0.00%) | 0.00 (0.00%) | 0.00 (0.00%) | 0.00 (0.00%) | 55.00 (7.61%) |
| *Lower Managerial and professional occupations* | 3.00 (7.32%) | 13.00 (21.31%) | 117.00 (82.39%) | 3.00 (3.53%) | 4.00 (5.48%) | 1.00 (0.79%) | 0.00 (0.00%) | 1.00 (0.96%) | 142.00 (19.64%) |
| *Intermediate occupations* | 0.00 (0.00%) | 0.00 (0.00%) | 3.00 (2.11%) | 73.00 (85.88%) | 0.00 (0.00%) | 2.00 (1.57%) | 2.00 (2.22%) | 2.00 (1.92%) | 82.00 (11.34%) |
| *Small employers and own account workers* | 0.00 (0.00%) | 0.00 (0.00%) | 11.00 (7.75%) | 1.00 (1.18%) | 60.00 (82.19%) | 0.00 (0.00%) | 0.00 (0.00%) | 0.00 (0.00%) | 72.00 (9.96%) |
| *Lower supervisory and technical occupations* | 0.00 (0.00%) | 0.00 (0.00%) | 0.00 (0.00%) | 0.00 (0.00%) | 1.00 (1.37%) | 118.00 (92.91%) | 2.00 (2.22%) | 4.00 (3.85%) | 125.00 (17.29%) |
| *Semi-routine occupations* | 0.00 (0.00%) | 0.00 (0.00%) | 5.00 (3.52%) | 4.00 (4.71%) | 0.00 (0.00%) | 0.00 (0.00%) | 79.00 (87.78%) | 1.00 (0.96%) | 89.00 (12.31%) |
| *Routine occupations* | 0.00 (0.00%) | 0.00 (0.00%) | 1.00 (0.70%) | 0.00 (0.00%) | 6.00 (8.22%) | 6.00 (4.72%) | 7.00 (7.78%) | 96.00 (92.31%) | 116.00 (16.04%) |
| N | 41.00 (5.67%) | 61.00 (8.44%) | 142.00 (19.64%) | 85.00 (11.76%) | 73.00 (10.10%) | 127.00 (17.57%) | 90.00 (12.45%) | 104.00 (14.38%) | 723.00 (100.00%) |
| Data Source BCS [Birth-Age30] | | | | | | | | | |

Table 3.6 Descriptive Statistics comparing RGSC by SOC2000 and SOC90 codes

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Semi-Dominant RGSC Social Class of Parents when Respondent was 10 SOC90 | | | | | | |
|  | *Professional* | *Managerial and Technical* | *Skilled non-manual* | *Skilled manual* | *Partly skilled* | *Unskilled* | Total |
| Semi-Dominant RGSC Social Class of Parents when Respondent was 10 SOC2000 |  |  |  |  |  |  |  |
| *Professional* | 39.00 (73.58%) | 1.00 (0.52%) | 0.00 (0.00%) | 0.00 (0.00%) | 0.00 (0.00%) | 0.00 (0.00%) | 40.00 (5.53%) |
| *Managerial and Technical* | 13.00 (24.53%) | 171.00 (89.06%) | 28.00 (23.93%) | 2.00 (0.85%) | 5.00 (4.81%) | 0.00 (0.00%) | 219.00 (30.29%) |
| *Skilled non-manual* | 0.00 (0.00%) | 8.00 (4.17%) | 79.00 (67.52%) | 1.00 (0.43%) | 1.00 (0.96%) | 0.00 (0.00%) | 89.00 (12.31%) |
| *Skilled manual* | 0.00 (0.00%) | 9.00 (4.69%) | 1.00 (0.85%) | 221.00 (94.44%) | 18.00 (17.31%) | 0.00 (0.00%) | 249.00 (34.44%) |
| *Partly skilled* | 1.00 (1.89%) | 3.00 (1.56%) | 4.00 (3.42%) | 8.00 (3.42%) | 75.00 (72.12%) | 3.00 (13.04%) | 94.00 (13.00%) |
| *Unskilled* | 0.00 (0.00%) | 0.00 (0.00%) | 5.00 (4.27%) | 2.00 (0.85%) | 5.00 (4.81%) | 20.00 (86.96%) | 32.00 (4.43%) |
| N | 53.00 (7.33%) | 192.00 (26.56%) | 117.00 (16.18%) | 234.00 (32.37%) | 104.00 (14.38%) | 23.00 (3.18%) | 723.00 (100.00%) |
| Data Source BCS [Birth-Age30] | | | | | | | |

Table 3.7 Descriptive Statistics comparing CAMSIS by SOC2000 and SOC90 codes

|  |  |
| --- | --- |
| CAMSIS2000 | |
| *Mean* | 49.71 |
| *Standard Deviation* | 13.82 |
| CAMSIS90 | |
| *Mean* | 50.13 |
| *Standard Deviation* | 14.53 |
| N | 723 |
| Data Source BCS [Birth-Age30] | |

## Modelling Main Economic Activity:

The primary outcome variable is the main economic activity of individuals in September of 1986. This is the first-month individuals were in when they received their O’level results after mandatory schooling. The first set of analyses estimates a multinomial logistic regression model. Table 2.6 details the deviance, change in deviance, change in degrees of freedom, and McFadden’s Adjusted Pseudo , AIC, and BIC measures to compare the null model with models of one explanatory variable. Table 2.7 details the exact statistics but through a sequential building of the null model with each subsequent independent variable added.

This model has been tested 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 and can be rejected at the 0.01 level.

The model output uses the reference category of education. The education category contrasts with all other economic activity categories because it has the most significant barrier to entry; continuing schooling expects previous educational merit. 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.

Table 3.8 Goodness-of-fit summaries for explanatory variables and Economic Activity

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s | AIC | BIC |
| Null Model | 1418.18 | - | - | - | 1424.18 | 1437.93 |
| Null Model + Educational Attainment | 1305.16 | 113.02 | 3 | 0.08 | 1317.16 | 1344.66 |
| Null Model + Sex | 1414.68 | 3.5 | 3 | 0.00 | 1426.68 | 1454.18 |
| Null Model + Tenure | 1403.27 | 14.91 | 3 | 0.01 | 1415.27 | 1442.77 |
| Null Model + NS-SEC (SOC 2000) | 1353.42 | 64.76 | 21 | 0.05 | 1401.42 | 1511.42 |

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

Table 3.9 Model building goodness-of-fit summaries for multinominal logistic regression model of Economic Activity

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 1418.18 | - | - | - | 1424.18 | 1437.93 |
| Null Model + Educational Attainment | 1305.16 | 113.02 | 3 | 0.08 | 1317.16 | 1344.66 |
| Null Model + Educational Attainment + Sex | 1299.29 | 5.87 | 3 | 0.08 | 1317.29 | 1358.54 |
| Null Model + Educational Attainment + Sex + Tenure | 1294.22 | 5.07 | 3 | 0.09 | 1318.22 | 1373.23 |
| Null Model + Educational Attainment + Sex + Tenure + NS-SEC (SOC 2000) | 1253.08 | 41.14 | 21 | 0.12 | 1319.08 | 1470.33 |

The model fit statistics demonstrate that there are normally distributed residuals and that the model is correctly specified. Table 2.7 suggests that deviance is reduced by 41.14 from the null for the full proposed model. AIC and BIC statistics suggest, unlike the NCDS model, that the full proposed model is not the best-fit model amongst those entered – however, the difference in such statistics is slight. Finally, the full model presents an adjusted pseudo-of 0.12. In other words, the full model explains 12 per cent of the variance of economic activity, leaving 88 per cent unexplained. This is a 12 per cent drop in compared to the NCDS model, suggesting that a temporal element of explanation may be necessary to the story of economic activity sorting. Ceteris paribus a 12 per cent drop in the explained variance of a model across time periods is suggestive of a substantively significant difference in the internal mechanisms that inform choice and opportunity for BCS youth in comparison to their NCDS peers at the same age. The following analysis with the full model is a complete records analysis with 723 observations.

As with the NCDS models, log odds will be presented alongside average marginal effects and quasi-variance statistics. On top of this, predicted probabilities, log odds and quasi-variance statistics are also graphed for a more intuitive understanding of the model.

The results of the multinomial logistic regression model are reported in Table 2.8. The output for employment demonstrates that individuals receiving five or more O’levels have decreased log odds of employment over education. Using average marginal effects there is a 12 per cent decreased probability for individuals to be employed over education if they received five or more O’levels. Sex and housing tenure had no statistically significant effect on an individual being in employment over being in education. NS-SEC also had no statistically significant effect on an individual's employment over education.

The output for training & apprenticeship category demonstrates that individuals that received five or more O’levels have a decreased log odds of being in training & apprenticeships compared to education. Using average marginal effects, there is a 17 per cent decreased probability for individuals to be in training & apprenticeships over education if they received five or more O’levels. Neither sex nor housing tenure. NS-SEC 4 and 5 are statistically significant – all other NS-SEC categories are statistically insignificant and will thus not be interpreted. Compared to NS-SEC 2, NS-SEC 4 has an increased log odds of being in training & apprenticeships compared to education. Translating this into average marginal effects demonstrates that there is a 11 per cent increased probability of individuals with a parental NS-SEC of 4 compared to 2 of being in training & apprenticeships over education. Similarly for those individuals from an NS-SEC 5 background compared to NS-SEC 2, there is an increased log odds of being in training & apprenticeships over education. Translated to average marginal effects, this results in an 11 per cent increased probability of being in training & apprenticeships over education for individuals in NS-SEC 5 compared to NS-SEC 2.

Finally, the unemployment & OLF category output demonstrates that individuals who received five or more O’levels have a decreased log odds of being in unemployment & OLF over education. Using average marginal effects, this translates to a 4 per cent decreased probability of being in unemployment & OLF category over education if individuals received five or more O’levels. Sex is statistically significant; men have an increased log odds of being unemployed & OLF over being in education compared to women. Translated to average marginal effects, men have a 2 per cent increased probability of being unemployed & OLF over education compared to women. Housing tenure is not statistically significant and so will not be interpreted. Beyond NS-SEC 4, no other NS-SEC category is statistically significant. Individuals that are in NS-SEC 4 compared to NS-SEC 2 have an increased log odds of being in unemployment & OLF over education. Translated to average marginal effects, this illustrates that being in NS-SEC 4 compared to NS-SEC 2 presents a 6 per cent increased probability of being in unemployment & OLF over education.

Table 3.10 Mlogit of Economic Activity

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NS-SEC | | | Average Marginal Effects | | Quasi-variance | | |
| Economic Activity: ‘School Reference Category | Coef. | S.E | Sig. | Prob. | S.E | S.E | LCI | UCI |
| Employment |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Five or More O’levels* | -1.37 | (0.24) | \*\*\* | -0.12 | (0.03) |  |  |  |
| Sex |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Male* | 0.33 | (0.22) |  | 0.03 | (0.03) |  |  |  |
| Housing Tenure |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Don't Own Home* | 0.25 | (0.26) |  | 0.02 | (0.03) |  |  |  |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |
| *1.1* | -1.32 | (0.78) |  | -0.12 | (0.05) | 0.74 | -2.83 | 0.18 |
| *1.2* | -0.54 | (0.54) |  | -0.06 | (0.06) | 0.48 | -1.53 | 0.45 |
| *2* | Ref. | (.) |  | (.) | (.) | 0.25 | -0.51 | 0.51 |
| *3* | -0.23 | (0.43) |  | -0.04 | (0.05) | 0.35 | -0.94 | 0.49 |
| *4* | 0.31 | (0.42) |  | -0.01 | (0.05) | 0.34 | -0.38 | 1.01 |
| *5* | 0.55 | (0.35) |  | 0.03 | (0.05) | 0.24 | 0.05 | 1.05 |
| *6* | 0.35 | (0.38) |  | 0.02 | (0.05) | 0.29 | -0.23 | 0.93 |
| *7* | 0.46 | (0.35) |  | 0.04 | (0.05) | 0.24 | -0.03 | 0.95 |
| Intercept | -1.14 | (0.28) | \*\*\* | (.) | (.) | (.) | (.) | (.) |
|  |  |  |  |  |  |  |  |  |
| School | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
|  |  |  |  |  |  |  |  |  |
| Training & Apprenticeships |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Five or More O’levels* | -1.87 | (0.28) | \*\*\* | -0.17 | (0.03) | (.) | (.) | (.) |
| Sex |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Male* | 0.27 | (0.23) |  | 0.01 | (0.03) | (.) | (.) | (.) |
| Housing Tenure |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Don't Own Home* | 0.36 | (0.26) |  | 0.04 | (0.03) | (.) | (.) | (.) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |
| *1.1* | 0.08 | (0.58) |  | 0.03 | (0.06) | 0.50 | -0.95 | 1.10 |
| *1.2* | -0.94 | (0.79) |  | -0.06 | (0.04) | 0.73 | -2.44 | 0.56 |
| *2* | Ref. | (.) |  | (.) | (.) | 0.31 | -0.63 | 0.63 |
| *3* | 0.41 | (0.45) |  | 0.05 | (0.05) | 0.33 | -0.27 | 1.09 |
| *4* | 1.09 | (0.44) | \* | 0.11 | (0.05) | 0.31 | 0.45 | 1.72 |
| *5* | 1.08 | (0.39) | \*\* | 0.11 | (0.05) | 0.24 | 0.59 | 1.58 |
| *6* | 0.72 | (0.43) |  | 0.07 | (0.05) | 0.29 | 0.13 | 1.31 |
| *7* | 0.63 | (0.40) |  | 0.05 | (0.04) | 0.26 | 0.10 | 1.16 |
| Intercept | -1.48 | (0.32) | \*\* | (.) | (.) | (.) | (.) |  |
|  |  |  |  |  |  |  |  |  |
| Unemployment & Out of Labour Force |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Five or More O’levels* | -3.18 | (1.04) | \*\* | -0.04 | (0.01) | (.) | (.) | (.) |
| Sex |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Male* | 1.03 | (0.48) | \* | 0.02 | (0.01) | (.) | (.) | (.) |
| Housing Tenure |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) | (.) |
| *Don't Own Home* | 0.26 | (0.54) |  | 0.00 | (0.01) | (.) | (.) | (.) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |
| *1.1* | -13.07 | (1070.37) |  | -0.01 | (0.01) | 1070.37 | -2199.06 | 2172.93 |
| *1.2* | 1.64 | (1.26) |  | 0.05 | (0.04) | 0.76 | 0.08 | 3.20 |
| *2* | Ref. | (.) |  | (.) | (.) | 1.02 | -2.07 | 2.07 |
| *3* | -12.79 | (759.94) |  | -0.01 | (0.01) | 759.94 | -1564.79 | 1539.21 |
| *4* | 2.42 | (1.13) | \* | 0.06 | (0.03) | 0.50 | 1.41 | 3.44 |
| *5* | 1.71 | (1.14) |  | 0.02 | (0.02) | 0.53 | 0.63 | 2.79 |
| *6* | 1.60 | (1.16) |  | 0.02 | (0.02) | 0.54 | 0.50 | 2.71 |
| *7* | 1.35 | (1.15) |  | 0.02 | (0.02) | 0.53 | 0.27 | 2.43 |
| Intercept | -4.23 | (1.05) | \*\*\* | (.) | (.) | (.) | (.) |  |
|  |  |  |  |  |  |  |  |  |
| Number of observations | 723 | | | | | | | |
| McFadden’s | 0.12 | | | | | | | |
| McFadden’s Adjusted Pseudo | 0.03 | | | | | | | |
| Cox-Snell Pseudo | 0.20 | | | | | | | |
| Nagelkerke Pseudo | 0.24 | | | | | | | |
| AIC | 1319.08 | | | | | | | |
| BIC | 1470.33 | | | | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: BCS [Birth-Age30]  Note: Complete Records Analysis | | | | | | | | |

Each variable is graphically visualised with its predicted probabilities to understand these results in a more manageable format. This allows for a more intuitive understanding of the multinominal logistic regression and provides a different outlook for interpretation. Each graph focuses on one variable within the model: educational attainment, sex, housing tenure, and NS-SEC, with each of the four economic activity outcome categories graphed using predicted probabilities.

A graph showing the number of probabilities

Description automatically generated

Figure 2.1 Predicted Probabilities of Economic Activity by NS-SEC

A graph showing the number of probabilities of economic activity

Description automatically generated

Figure 2.2 Predicted Probabilities of Economic Activity by Educational Attainment

A graph showing the number of individuals in the same age

Description automatically generated with medium confidence

Figure 2.3 Predicted Probabilities of Economic Activity by Sex

A graph showing the number of individuals in the economic activity

Description automatically generated with medium confidence

Figure 2.4 Predicted Probabilities of Economic Activity by Housing Tenure

Alongside the graphical presentation of predicted probabilities, the following figures also visualise the log odds of NS-SEC within each outcome category (except the reference category of education) alongside quasi-variance statistics to overcome the reference category problem. The reference category within NS-SEC is NS-SEC 2 which is demonstrated by the lack of confidence intervals at NS-SEC 2 for each graph.

A graph of a graph showing the number of individuals

Description automatically generated with medium confidence

Figure 2.5 Log Odds versus Quasi-variance statistics of individuals being in Education over Employment

A graph showing the number of numbers and the number of logistic

Description automatically generated with medium confidence

Figure 2.6 Log Odds versus Quasi-variance statistics of individuals being in Education over Training & Apprenticeships

A graph with red and black lines

Description automatically generated

Figure 2.7 Log Odds versus Quasi-variance statistics of individuals being in Education over Unemployment & OLF

### Discussion and Conclusion

The multinomial logistic regression model indicates that some structural inequalities impact an individual’s choice of sorting into economic activity post-mandatory schooling. Educational attainment was statistically and substantively significant across all economic activity categories. Across both employment and training & apprenticeship outcome categories, educational attainment was by far the single largest substantive impact upon an individual’s sorting into said economic activity over education.

Whilst NS-SEC has had a comparative decline in influence upon an individual’s trajectory into economic activity post-mandatory schooling compared to the NCDS model, NS-SEC still has a substantive impact on individual choice and opportunity – though this is dependent on the particular outcome category. Whilst there is no NS-SEC based impact on an individual’s decision to be in employment over education, there are substantially significant class based impacts on an individual’s pathways into training & apprenticeships as well as unemployment & OLF. The NS-SEC categories that are found to be significant within the model are indicative of a particular effect for those individuals that resided within small employers and own account workers and lower supervisor and technical occupations origins compared to lower managerial, administrative and professional occupations. An alternative social stratification schema such as RGSC with its focus on a stricter manual/non-manual divide or CAMSIS with its metric nature may provide alternative findings. A definitive conclusion would thus be premature prior to conducting a full sensitivity analysis of social stratification measures.

Sex as a variable was only found to be statistically significant across only one of the outcomes of economic activity. Even then, the substantive effect of sex is minuet – at only 2 per cent. Compared to NCDS models it appears that sex has had a declined impact upon individuals sorting into economic activity post-mandatory schooling.

Housing tenure as a form of structural inequality was not found to be statistically significant across the given model. This initially may suggest that there is no substantive impact that housing tenure holds upon individuals from the BCS cohort entering into economic activity post-mandatory schooling. This finding would stand in stark contrast to the new structuralist thesis – promoted by the likes of Saunders et al (Saunders, 2003, 2021) that have suggested the decline in traditional structural inequalities such as social class and sex is a result of an increasing influence of housing tenure. Whilst this may be an attractive finding to promote initially, on further inspection of the data within the model there are certain discrepancies that need to be addressed prior to this conclusion being made. The single largest discrepancy originates from the level of missingness within the overall model. Over half of the potential observations within the model are wiped out due to item missingness primarily at the educational attainment variable. This missingness does appear to skew the data within the model when comparing it to representative data from the same time period (HomeOwners Alliance, 2012). The model proposed as an inflated level of homeowners compared to non-home owners as well as an inflated number of women in comparison to men. Until a robust handling of missing data occurs, no firm conclusion can be made concerning these findings.

## Sensitivity Analysis of Independent Variables

Following the NCDS chapter, this section seeks to present a sensitivity analysis of social stratification measures to provide an informed assessment of which social stratification measure to use within the given BCS model. As with the NCDS, NS-SEC, CAMSIS, and RGSC are the three measures that will be used within subsequent sensitivity analysis. Following the NS-SEC model using BCS data there appears to be a significant class effect for individuals within certain class origins with selected economic activity outcomes. The task of this next chapter is to explore if this phenomena exists across social stratification measures or if different measures of parental social stratification demonstrate unique and distinct impacts upon individuals sorting into economic activity outcomes post-mandatory schooling.

### Testing Measures of Parental Social Class

Three separate multinomial logistic regressions are presented in Table 2.9. The first model has been described at length in the previous section and uses NS-SEC. The second model uses CAMSIS, and the third model uses RGSC. These models are all presented using log odds and average marginal effects to enhance interpretation and comparison between models. Substantive interpretation will only be made on variables within models that have achieved statistical significance.

Educational attainment is substantively identical across all models, including the NS-SEC, CAMSIS, and RGSC models. All three models also present substantively identical results for sex at the unemployment & OLF. The CAMSIS model presents statistically significant results for CAMSIS across employment and training & apprenticeship categories. There is no per cent change in the models when looking at average marginal effects suggesting that if a CAMSIS model is adopted, there is an indication that parental social stratification measured through CAMSIS does not have a substantive impact upon individuals sorting into economic activity post-mandatory schooling.

The RGSC model diverges the most from the NS-SEC model discussed earlier. Whilst the NS-SEC model does not present any statistically significant results within the employment outcome category, RGSC does. Both RGSC 3M and 4 are statistically significant. Compared to RGSC 2, RGSC 3M sees individuals have an increased log odds of being in employment over education. Translated into average marginal effects, this results in a 6 per cent increased probability of individuals with RGSC 3M origins compared to RGSC 2 origins of being in employment over education. Compared to RGSC 2, RGSC 4 also sees individuals have an increased log odds of being in employment over education – in average marginal effects, this translates to a 7 per cent increased probability of being in employment over education. The RGSC model thus demonstrates a substantive difference between the NS-SEC model. The former model establishes that there is a distinct and substantive manual/non-manual divide amongst individuals entering employment over education post-mandatory education. The training & apprenticeship category also offers some minor deviations between RGSC and NS-SEC models. Whilst both demonstrate that there are statistically significant categories (NS-SEC 4 and 5 & RGSC 3M and 4) the substantive significance differs between models. Whilst NS-SEC 4 and 5 have a 11 per cent increased probability of being in training & apprenticeships over education compared to individuals in NS-SEC 2, the RGSC model presents results that RGSC 3M and 4 have a 6 and 7 per cent increased probability of being in training & apprenticeships over education compared to individuals in RGSC 2 respectively. When comparing the differences between categories, NS-SEC 4 and 5 (Small employers and own account workers & lower supervisory and technical occupations) generally line up with RGSC 3M and 4 (Skilled manual occupations & partly skilled occupations). The substantive difference between models – of around 5 per cent – can most likely be characterised as slight differences in the way each social stratification measure is constructed. Whilst there is a substantive difference between the models, the general positive trend is the same. Finally, the unemployed & OLF category also sees a slight deviation between NS-SEC and RGSC models. Whilst NS-SEC 4 and RGSC 3M share substantively identical interpretations, the RGSC model finds – unlike the NS-SEC model – that there is a substantive difference between those that have social origins in the top of the social stratification schema compared to the reference category. The RGSC model demonstrates that those individuals in RGSC 1 compared to RGSC 2 have an increased log odds of being in unemployment & OLF over education. Translated to average marginal effects, this results in a 7 per cent increased probability of individuals from RGSC 1 social origins of being in unemployment & OLF over education compared to individuals from RGSC 2 social origins. This is a finding not duplicated across either NS-SEC nor CAMSIS models. The findings within the unemployment & OLF category as mentioned previously may have missing data implications. Within the RGSC model RGSC 3NM has a standard error of 737.60 which is almost entirely related to the low observations within that category. Combined with an early discussion of the implications that missing data at educational attainment has had on the distribution of observations within the given models, missing data may have a potentially serious implication regarding the results of these models.

The goodness-of-fit statistics for all three models is reported within Table 2.9. Differences in measures exist, but the minor nature of these differences indicates the amount of variance explained across the three models remains consistent. AIC statistics across all three models is stable, with minimal differences across the models. Overall, the CAMSIS model using AIC along appears to be the best fit model – this is not surprising considering that AIC and BIC statistics tend to favour metric over categorical measures within models. This best fit finding is replicated using BIC statistics, though the differences between the BIC across models is much larger compared to AIC statistics. Given the fact that AIC and BIC favour metric based measures, and that the difference between CAMSIS and RGSC models with respect to said goodness-of-fit statistics is marginal, the RGSC model is selected for further investigation going forward.

Table 3.11 Sensitivity analyses of alternative measures of parental social stratification

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NS-SEC | | | | | CAMSIS | | | | | RGSC | | | | |
| Economic Activity: ‘School Reference Category | Coef. | S.E | Sig. | Prob. | S.E | Coef. | S.E | Sig. | Prob. | S.E | Coef. | S.E | Sig. | Prob. | S.E |
| Employment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -1.37 | (0.24) | \*\*\* | -0.12 | (0.03) | -1.33 | (0.24) | \*\*\* | -0.12 | (0.03) | -1.38 | (0.24) | \*\*\* | -0.12 | (0.03) |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) |  |  |  | (.) | (.) |  |  |  | (.) | (.) |
| *Male* | 0.33 | (0.22) |  | 0.03 | (0.03) | 0.31 | (0.22) |  | 0.03 | (0.03) | 0.30 | (0.22) |  | 0.02 | (0.03) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) |  |  |  | (.) | (.) |  |  |  | (.) | (.) |
| *Don't Own Home* | 0.25 | (0.26) |  | 0.02 | (0.03) | 0.25 | (0.25) |  | 0.02 | (0.03) | 0.30 | (0.26) |  | 0.03 | (0.03) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1.1* | -1.32 | (0.78) |  | -0.12 | (0.05) |  |  |  |  |  |  |  |  |  |  |
| *1.2* | -0.54 | (0.54) |  | -0.06 | (0.06) |  |  |  |  |  |  |  |  |  |  |
| *2* | Ref. | (.) |  | (.) | (.) |  |  |  |  |  |  |  |  |  |  |
| *3* | -0.23 | (0.43) |  | -0.04 | (0.05) |  |  |  |  |  |  |  |  |  |  |
| *4* | 0.31 | (0.42) |  | -0.01 | (0.05) |  |  |  |  |  |  |  |  |  |  |
| *5* | 0.55 | (0.35) |  | 0.03 | (0.05) |  |  |  |  |  |  |  |  |  |  |
| *6* | 0.35 | (0.38) |  | 0.02 | (0.05) |  |  |  |  |  |  |  |  |  |  |
| *7* | 0.46 | (0.35) |  | 0.04 | (0.05) |  |  |  |  |  |  |  |  |  |  |
| CAMSIS (SOC 2000) |  |  |  |  |  | -0.02 | (0.01) | \*\* | -0.00 | (0.00) |  |  |  |  |  |
| RGSC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1* |  |  |  |  |  |  |  |  |  |  | -0.20 | (0.58) |  | -0.02 | (0.06) |
| *2* |  |  |  |  |  |  |  |  |  |  | Ref. | (.) |  | (.) | (.) |
| *3NM* |  |  |  |  |  |  |  |  |  |  | -0.19 | (0.41) |  | -0.03 | (0.04) |
| *3M* |  |  |  |  |  |  |  |  |  |  | 0.67 | (0.27) | \* | 0.06 | (0.04) |
| *4* |  |  |  |  |  |  |  |  |  |  | 0.68 | (0.35) | \* | 0.07 | (0.05) |
| *5* |  |  |  |  |  |  |  |  |  |  | 0.40 | (0.54) |  | 0.02 | (0.07) |
| Intercept | -1.14 | (0.28) | \*\*\* | (.) | (.) | 0.15 | (0.43) |  |  |  | -1.32 | (0.25) | \*\*\* |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School |  |  |  |  |  | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Training & Apprenticeships |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -1.87 | (0.28) | \*\*\* | -0.17 | (0.03) | -1.80 | (0.28) | \*\*\* | -0.16 | (0.03) | -1.85 | (0.28) | \*\*\* | -0.17 | (0.03) |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 0.27 | (0.23) |  | 0.01 | (0.03) | 0.25 | (0.22) |  | 0.02 | (0.03) | 0.27 | (0.22) |  | 0.02 | (0.03) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Don't Own Home* | 0.36 | (0.26) |  | 0.04 | (0.03) | 0.36 | (0.25) |  | 0.04 | (0.03) | 0.35 | (0.26) |  | 0.03 | (0.03) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1.1* | 0.08 | (0.58) |  | 0.03 | (0.06) | (.) | (.) |  |  |  | (.) | (.) |  |  |  |
| *1.2* | -0.94 | (0.79) |  | -0.06 | (0.04) |  |  |  |  |  |  |  |  |  |  |
| *2* | Ref. | (.) |  | (.) | (.) |  |  |  |  |  |  |  |  |  |  |
| *3* | 0.41 | (0.45) |  | 0.05 | (0.05) |  |  |  |  |  |  |  |  |  |  |
| *4* | 1.09 | (0.44) | \* | 0.11 | (0.05) |  |  |  |  |  |  |  |  |  |  |
| *5* | 1.08 | (0.39) | \*\* | 0.11 | (0.05) |  |  |  |  |  |  |  |  |  |  |
| *6* | 0.72 | (0.43) |  | 0.07 | (0.05) |  |  |  |  |  |  |  |  |  |  |
| *7* | 0.63 | (0.40) |  | 0.05 | (0.04) |  |  |  |  |  |  |  |  |  |  |
| CAMSIS (SOC 2000) |  |  |  |  |  | -0.03 | (0.01) | \*\* | -0.00 | (0.00) |  |  |  |  |  |
| RGSC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1* |  |  |  |  |  | (.) | (.) |  |  |  | -0.67 | (0.78) |  | -0.06 | (0.05) |
| *2* |  |  |  |  |  |  |  |  |  |  | Ref. | (.) |  | (.) | (.) |
| *3NM* |  |  |  |  |  |  |  |  |  |  | 0.44 | (0.38) |  | 0.06 | (0.05) |
| *3M* |  |  |  |  |  |  |  |  |  |  | 0.75 | (0.29) | \* | 0.06 | (0.03) |
| *4* |  |  |  |  |  |  |  |  |  |  | 0.76 | (0.36) | \* | 0.07 | (0.04) |
| *5* |  |  |  |  |  |  |  |  |  |  | 0.80 | (0.52) |  | 0.08 | (0.07) |
| Intercept | -1.48 | (0.32) | \*\* | (.) | (.) | 0.34 | (0.45) |  |  |  | -1.39 | (0.26) | \*\*\* |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Unemployment & Out of Labour Force |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.18 | (1.04) | \*\* | -0.04 | (0.01) | -3.04 | (1.04) | \*\* | -0.04 | (0.01) | -3.18 | (1.04) | \*\* | -0.04 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 1.03 | (0.48) | \* | 0.02 | (0.01) | 0.98 | (0.48) | \* | 0.02 | (0.01) | 1.02 | (0.48) | \* | 0.02 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Don't Own Home* | 0.26 | (0.54) |  | 0.00 | (0.01) | 0.18 | (0.52) |  | 0.00 | (0.01) | 0.44 | (0.54) |  | 0.01 | (0.01) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1.1* | -13.07 | (1070.37) |  | -0.01 | (0.01) | (.) | (.) |  |  |  | (.) | (.) |  |  |  |
| *1.2* | 1.64 | (1.26) |  | 0.05 | (0.04) |  |  |  |  |  |  |  |  |  |  |
| *2* | Ref. | (.) |  | (.) | (.) |  |  |  |  |  |  |  |  |  |  |
| *3* | -12.79 | (759.94) |  | -0.01 | (0.01) |  |  |  |  |  |  |  |  |  |  |
| *4* | 2.42 | (1.13) | \* | 0.06 | (0.03) |  |  |  |  |  |  |  |  |  |  |
| *5* | 1.71 | (1.14) |  | 0.02 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| *6* | 1.60 | (1.16) |  | 0.02 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| *7* | 1.35 | (1.15) |  | 0.02 | (0.02) |  |  |  |  |  |  |  |  |  |  |
| CAMSIS (SOC 2000) |  |  |  |  |  | -0.04 | (0.02) | \* | -0.00 | (0.00) |  |  |  |  |  |
| RGSC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1* |  |  |  |  |  |  |  |  |  |  | 2.57 | (1.27) | \* | 0.07 | (0.05) |
| *2* |  |  |  |  |  |  |  |  |  |  | Ref. | (.) |  | (.) | (.) |
| *3NM* |  |  |  |  |  |  |  |  |  |  | -12.40 | (737.60) |  | -0.01 | (0.01) |
| *3M* |  |  |  |  |  |  |  |  |  |  | 2.63 | (1.05) | \* | 0.05 | (0.01) |
| *4* |  |  |  |  |  |  |  |  |  |  | 1.44 | (1.25) |  | 0.01 | (0.01) |
| *5* |  |  |  |  |  |  |  |  |  |  | 1.80 | (1.47) |  | 0.02 | (0.03) |
| Intercept | -4.23 | (1.05) | \*\*\* | (.) |  | -0.99 | (0.98) |  |  |  | -4.79 | (1.06) | \*\*\* |  |  |
| Number of observations | 723 | | | | | 723 | | | | | 723 | | | | |
| McFadden’s | 0.12 | | | | | 0.10 | | | | | 0.11 | | | | |
| McFadden’s Pseudo | 0.03 | | | | | 0.05 | | | | | 0.04 | | | | |
| Cox-Snell Pseudo | 0.20 | | | | | 0.18 | | | | | 0.20 | | | | |
| Nagelkerke Pseudo | 0.24 | | | | | 0.21 | | | | | 0.23 | | | | |
| AIC | 1319.08 | | | | | 1308.04 | | | | | 1312.82 | | | | |
| BIC | 1470.33 | | | | | 1376.79 | | | | | 1436.57 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: BCS [Birth-Age 30]  Note: Complete Records Analysis for NS-SEC, CAMSIS, RGSC | | | | | | | | | | | | | | | |

### Discussion and Conclusions

There are three substantive findings from this sensitivity analysis. Firstly, the CAMSIS model whilst at first glance the overall best fit when looking at AIC and BIC statistics, in actuality is more likely put down to a bias within these statistics to favour metric based variables over categorical ones. This is likely to be the correct assumption due to the overall minor differences between AIC and BIC statistics between the models. This being said, the CAMSIS model does offer a different interpretation on the substantive effect of individuals parental social origins and their impact upon an individual’s sorting into economic activity post-mandatory schooling. The CAMSIS model indicates that there is zero substantive impact that social origins has upon and individuals first instance of economic activity. Whilst this is interesting, given the marginal AIC and BIC statistics as well as the bias mentioned above, CAMSIS is not the best model fit going forward. The second substantive finding to come from this sensitivity analysis is that NS-SEC as a model is also not the best fit going forward. Whilst presenting interesting substantive findings, it also boasts as a model the largest AIC and BIC statistics of the three models. The third substantive finding relates to the RGSC model being the best model fit going forward. This relates to its goodness-of-fit statistics. On top of this, when comparing the RGSC model with the NS-SEC model there are some critical substantive differences.

Going forward, missingness will be explored in greater depth using the RGSC model. Throughout model interpretation missing data and the level of missingness have come up as possible explanations for the substantive interpretation of the model – as well as explanations for the large standard errors. Multiple imputation of the RGSC model will compare the substantive interpretations of an imputed model versus a complete records analysis. Given the greater than 50 per cent missingness within the model alternative solutions will also be presented.

## Sensitivity analysis using SOC codes

### SOC Codes Modelling

Table 3.12 Goodness-of-fit summaries for explanatory variables and Economic Activity Comparing SOC codes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 1418.18 | - | - | - | 1424.18 | 1437.93 |
| Null Model + Educational Attainment | 1305.16 | 113.02 | 3 | 0.08 | 1317.16 | 1344.66 |
| Null Model + Sex | 1414.68 | 3.5 | 3 | 0.00 | 1426.68 | 1454.18 |
| Null Model + Tenure | 1403.27 | 14.91 | 3 | 0.01 | 1415.27 | 1442.77 |
| Null Model + RGSC (SOC 2000) | 1362.61 | 55.57 | 15 | 0.04 | 1398.61 | 1481.11 |
| Null Model + RGSC (SOC 90) | 1364.67 | 53.51 | 15 | 0.04 | 1400.67 | 1483.17 |

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

Table 3.13 Model building goodness-of-fit summaries for multinominal logistic regression model of Economic Activity Comparing SOC codes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 1418.18 | - | - | - | 1424.18 | 1437.93 |
| Null Model + Educational Attainment | 1305.16 | 113.02 | 3 | 0.08 | 1317.16 | 1344.66 |
| Null Model + Educational Attainment + Sex | 1299.29 | 5.87 | 3 | 0.08 | 1317.29 | 1358.54 |
| Null Model + Educational Attainment + Sex + Tenure | 1294.22 | 5.07 | 3 | 0.09 | 1318.22 | 1373.23 |
| Null Model + Educational Attainment + Sex + Tenure + RGSC (SOC 2000) | 1258.82 | 34.40 | 15 | 0.11 | 1312.82 | 1436.57 |
| Null Model + Educational Attainment + Sex + Tenure + RGSC (SOC 90) | 1260.89 | 33.33 | 15 | 0.11 | 1314.89 | 1438.65 |

Table 3.14 Sensitivity Analysis of SOC Codes

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SOC 2000 Codes | | | | | SOC 90 Codes | | | | |
|  | RGSC | | | Average Marginal Effects | | RGSC | | | Average Marginal Effects | |
| Economic Activity: ‘School Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** |
| Employment |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -1.38 | (0.24) | \*\*\* | -0.12 | (0.03) | -1.37 | (0.24) | \*\*\* | -0.12 | (0.03) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 0.30 | (0.22) |  | 0.02 | (0.03) | 0.30 | (0.22) |  | 0.02 | (0.03) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.30 | (0.26) |  | 0.03 | (0.03) | 0.27 | (0.26) |  | 0.02 | (0.03) |
| RGSC |  |  |  |  |  |  |  |  |  |  |
| *1* | -0.20 | (0.58) |  | -0.02 | (0.06) | -0.21 | (0.53) |  | -0.02 | (0.05) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3NM* | -0.19 | (0.41) |  | -0.03 | (0.04) | -0.07 | (0.37) |  | -0.01 | (0.04) |
| *4M* | 0.67 | (0.27) | \* | 0.06 | (0.04) | 0.59 | (0.29) | \* | 0.06 | (0.04) |
| *4* | 0.68 | (0.35) | \* | 0.07 | (0.05) | 0.79 | (0.35) | \* | 0.08 | (0.05) |
| *5* | 0.40 | (0.54) |  | 0.02 | (0.07) | 0.95 | (0.58) |  | 0.10 | (0.09) |
| Intercept | -1.32 | (0.25) | \*\*\* | (.) | (.) | -1.32 | (0.26) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| School | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Training & Apprenticeships |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -1.85 | (0.28) | \*\*\* | -0.17 | (0.03) | -1.84 | (0.28) | \*\*\* | -0.17 | (0.03) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 0.27 | (0.22) |  | 0.02 | (0.03) | 0.24 | (0.22) |  | 0.01 | (0.03) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.35 | (0.26) |  | 0.03 | (0.03) | 0.36 | (0.26) |  | 0.03 | (0.03) |
| RGSC |  |  |  |  |  |  |  |  |  |  |
| *1* | -0.67 | (0.78) |  | -0.06 | (0.05) | -0.35 | (0.59) |  | -0.04 | (0.05) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3NM* | 0.44 | (0.38) |  | 0.06 | (0.05) | 0.05 | (0.37) |  | 0.01 | (0.04) |
| *4M* | 0.75 | (0.29) | \* | 0.06 | (0.03) | 0.39 | (0.30) |  | 0.02 | (0.03) |
| *4* | 0.76 | (0.36) | \* | 0.07 | (0.04) | 0.89 | (0.35) | \* | 0.09 | (0.05) |
| *5* | 0.80 | (0.52) |  | 0.08 | (0.07) | 0.93 | (0.60) |  | 0.09 | (0.09) |
| Intercept | -1.39 | (0.26) | \*\*\* | (.) | (.) | -1.23 | (0.27) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Unemployment & Out of Labour Force |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.18 | (1.04) | \*\* | -0.04 | (0.01) | -3.25 | (1.04) | \*\* | -0.04 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 1.02 | (0.48) | \* | 0.02 | (0.01) | 1.02 | (0.48) | \* | 0.02 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.44 | (0.54) |  | 0.01 | (0.01) | 0.62 | (0.54) |  | 0.01 | (0.02) |
| RGSC |  |  |  |  |  |  |  |  |  |  |
| *1* | 2.57 | (1.27) | \* | 0.07 | (0.05) | 1.11 | (0.96) |  | 0.04 | (0.04) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3NM* | -12.40 | (737.60) |  | -0.01 | (0.01) | -13.70 | (613.77) |  | -0.02 | (0.01) |
| *4M* | 2.63 | (1.05) | \* | 0.05 | (0.01) | 1.28 | (0.67) |  | 0.03 | (0.02) |
| *4* | 1.44 | (1.25) |  | 0.01 | (0.01) | 0.07 | (0.96) |  | -0.01 | (0.02) |
| *5* | 1.80 | (1.47) |  | 0.02 | (0.03) | -13.72 | (1558.62) |  | -0.02 | (0.01) |
| Intercept | -4.79 | (1.06) | \*\*\* | (.) | (.) | -3.55 | (0.68) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Number of observations | 723 | | | | | 723 | | | | |
| McFadden’s | 0.11 | | | | | 0.11 | | | | |
| McFadden’s Adjusted Pseudo | 0.04 | | | | | 0.04 | | | | |
| Cox-Snell Pseudo | 0.20 | | | | | 0.20 | | | | |
| Nagelkerke Pseudo | 0.23 | | | | | 0.23 | | | | |
| AIC | 1312.82 | | | | | 1314.89 | | | | |
| BIC | 1436.57 | | | | | 1438.65 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: BCS [Birth-Age 30]  Note: Complete Records Analysis | | | | | | | | | | |

### Discussion and Conclusion

## Missing Data in the BCS

There are 1,616 individuals identified in the BCS who indicated some form of economic activity within the 21-sub-sample sweep. With all variables added to the model, there are 958 observations with missing data on at least one of the variables included for analysis – the model itself is 44 per cent complete, according to Table 2.10. Using Table 2.11, it can be illustrated that of the missingness amongst the variables, 29 were missing in economic activity, 805 were missing at educational attainment, 85 were missing at housing tenure, 0 were missing at sex, and 251 were missing at RGSC. The most considerable missingness can be attributed to the educational attainment variable. This is primarily because of a failure of the BCS survey to ask all participants to answer the relevant educational attainment questions – and a lack of follow-up in further sweeps.

Patterns of missing data are presented in Table 2.10. Within the BCS sample, 44 per cent have complete records on all variables, 39 per cent are missing values at educational attainment, and a further 5 per cent were missing at both educational attainment and RGSC. Finally, 4 per cent were missing at solely RGSC. All other missingness is at 3 per cent or lower.

Educational attainment is the only variable within the model that takes data from individuals post-16 years of age. This is important considering that pre-26 year sweep cohort members were followed up by parental interview and examination. From age 26 onwards the BCS converted to a postal questionnaire format that was sent to cohort members (Elliott and Shepherd, 2006). The age 26 sweep only maintains 55 per cent of the original sample according to Table 2.2. This is in large part due to a failure to move a more sustainable and stable track and trace system for cohort members after leaving school and the childhood home[[23]](#footnote-23). The fact that the educational attainment variable is taken from data post-16 means that it has a higher amount of missingness compared to all other variables that are taken closer to birth, or taken prior to the substantial loss of a number of cohort members.

Table 3.15 Missing data patterns for BCS

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N | Percent Complete (%) | Educational Attainment | Economic Activity | Housing Tenure | RGSC | Sex |
| 723 | 44 | ✓ | ✓ | ✓ | ✓ | ✓ |
| 645 | 39 |  | ✓ | ✓ | ✓ | ✓ |
| 84 | 5 |  | ✓ | ✓ |  | ✓ |
| 60 | 4 | ✓ | ✓ | ✓ |  | ✓ |
| Total = 1645 |

Table 3.16 Number of Observations missing for BCS

|  |  |  |
| --- | --- | --- |
| Outcome Variable: Economic Activity | Obs=. | Obs<. |
| Economic Activity | 29 | 1616 |
| Educational Attainment | 805 | 840 |
| Housing Tenure | 85 | 1560 |
| Sex | 0 | 1645 |
| RGSC | 251 | 1394 |

Given that a complete records analysis can only be undertaken if data is confidently considered to be MCAR, the patterns of missingness related to this model suggest that data may be MAR; educational attainment has already been discussed, but looking more closely at the distribution of other variables within the model suggests missingness may have an impact upon the substantive interpretation of results. Sex, housing tenure, and RGSC. A typical solution to a model that appears to fit an MAR assumption would be to conduct some form of multiple imputation and compare the substantive findings of the model. The BCS model has 56 per cent missingness. Whilst some literature on missing data advocates for the use of imputation on data even as high as 90 per cent missingness (Madley-Dowd *et al.*, 2019), others argue that 50 per cent is a more acceptable threshold (Carpenter and Kenward, 2012). Contemporary literature on the topic has investigated simulations on missing data and concludes that multiple imputation can be used on large amounts of missing data within a model so long as the imputation model (and its auxiliary variables) are robust however large amounts of missing data on a model does present a large amount of random variance (Hardt *et al.*, 2013). Due to the possible ramifications of variance upon an imputed model with large amounts of missingness, two other models will also be produced to compare their substantive effects alongside the imputed and complete records analysis. The first model will seek to recode educational attainment in a ‘worst case scenario’. This model will recode all missingness at educational attainment equal to zero. In other words all observations that are missing at educational attainment were coded as individuals receiving less than five O’levels. The second model is a ‘best case scenario' model. This recodes all missingness at educational attainment to one. In other words all observations that are missing at educational attainment were coded as individuals receiving five or more O’levels. Placing these two models alongside a complete records analysis as well as an imputed model will accomplish two aims. The first, is to aid in the MAR assumption – if variables such as housing tenure, sex, and RGSC have been skewed in their substantive interpretation because of the amount of missingness present at the complete records analysis then these two models alongside the imputation model will aid in this interpretation. The second, is to compare these two models with the imputation model directly, comparing their substantive interpretations across the three models to see if it is more likely that the bulk of missingness at educational attainment came from those individuals that received less than five or five or more O’levels.

Prior to model interpretation, imputation was conducted using 60 imputations at a burnin of 20[[24]](#footnote-24). Trace plots were produced for each analytical variable within the model, with the exception of sex – which does not have any level of missingness and thus did not need to be imputed. These trace plots can be seen from figures 2.8-11.

A graph showing the number of numbers

Description automatically generated with medium confidence

Figure 2.8 Trace plot summaries for Economic Activity

A graph of different colored lines

Description automatically generated with medium confidence

Figure 2.9 Trace plot summaries for Educational Attainment

A graph of a number of numbers

Description automatically generated with medium confidence

Figure 2.10 Trace plot summaries for RGSC

A graph showing different colored lines

Description automatically generated with medium confidence

Figure 2.11 Trace plot summaries for Housing Tenure

Model interpretation is pulled from Table 2.12. The substantive interpretation of the complete records analysis has already been undertaken. The following interpretation will thus focus upon the two models where missingness at educational attainment has been recoded to equal zero or one, as well as the multiple imputation model. Reference will be made to the complete records analysis when appropriate. Interpretation of the models will proceed by going through each variable within each category of the outcome variable, stating its statistical significant, substantive significance, and finally its deviation from other models. Finally a summation will be given on the four models and the most appropriate model fit going forward – goodness-of-fit statistics like AIC and BIC are not able to be used with imputed models, thus a substantive judgement will be made instead.

Starting with the employment outcome in reference to education three models find educational attainment to be statistically significant – with the model that codes all educational attainment missingness equal to one being the only one that does not find it to be statistically significant. Across the statistically significant models, the substantive interpretations however differ. Whilst the complete records analysis shows that those individuals that received five or more O’levels compared to those that did not were 12 per cent less likely to enter into employment over education post-mandatory schooling. Comparing this to the other models and we see for the model where all educational attainment missingness is coded as zero the average marginal effect rises to 17 per cent decreased probability and for the imputed model it decreases to 10 per cent decreased probability. Moving on to sex, whilst the complete records analysis finds sex not to be statistically significant, all other models do. All models show that there is an increased log odds, translated to a 7 per cent increased probability of men entering into employment over education compared to women. Moving on to housing tenure, whilst the complete records analysis finds housing tenure not to be statistically significant, all other models do. Whilst all other models find those that don’t own their own home to have an increased log odds of being in employment over education. When translated to average marginal effects there is a minor deviation between the models that amounts to a 3 per cent difference between the three models. Moving on to RGSC, all four models find RGSC 3M and 4 statistically significant. All four models show that individuals with a RGSC 3M or 4 social origins have an increased log odds of being in employment over education compared to their RGSC 2 counterparts. For RGSC 3M there is minor deviation across the models resulting in a 4 per cent difference when translated into average marginal effects. For RGSC 4 there is also a minor deviation across the models resulting in a 4 per cent difference when translated into average marginal effects.

Looking now at the training & apprenticeship category in reference to education, three models find educational attainment to be statistically significant – with the model that codes all educational attainment missingness equal to one being the only one that does not find it to be statistically significant. The three models that find educational attainment to be statistically significant all find individuals with five or more O’levels to have a decreased log odds of being in training & apprenticeships over education compared to individuals with less than five O’levels. Across the statistically significant models the substantive interpretation does differ however. Whilst the complete records analysis model reports a 17 per cent decreased probability of individuals that receive five or more O’levels of being in training & apprenticeships over education compared to those individuals that received less than five O’levels. The model that recoded all educational attainment missingness equal to zero reported an average marginal effect of a decreased probability of 24 per cent. The imputed model reports an average marginal effect of a decreased probability of 22 per cent. Moving on to sex, whilst the complete records analysis finds sex not to be statistically significant, all other models do. All models show that there is an increased log odds, translated to 2 per cent for both the models that recoded all educational attainment missingness equal to zero and to one and 3 per cent for the imputed model. Moving on to housing tenure, whilst the complete records analysis finds housing tenure not to be statistically significant, all other models do. Whilst all other models find those that didn’t own their own homes to have an increased log odds of being in training & apprenticeships over education compared to those that do own their own home. In relation to RGSC, both RGSC 3M and 4 are statistically significant across all four models. For RGSC 3M there is a small difference in average marginal effects across all four models – around 5 per cent. For RGSC 4 there is again a small difference in average marginal effects across the four models – around 5 per cent. Whilst the complete records analysis does not find any other RGSC categories as statistically significant, all other models find RGSC 5 to be statistically significant. For the model that has recoded all missingness at educational attainment equal to zero, those individuals that have RGSC 5 origins have a 10 per cent increased probability of being in training & apprenticeships over education compared to those from RGSC 2 origins. For the model that has recoded all missingness at educational attainment equal to one, this results in a 15 per cent increased probability. Finally, for the imputation model, this results in a 11 per cent increased probability. With respect to the models that share the most in common, the model that recodes all missingness at educational attainment equal to zero and the imputation model share the most similar substantive interpretation.

Moving on the unemployment & OLF category in reference to education, three models find educational attainment to be statistically significant – with the model that codes all educational attainment missingness equal to one being the only model that does not find it to be statistically significant. The three models that find educational attainment to be statistically significant all find individuals with five or more O’levels to have a decreased log odds of being in unemployment & OLF over education compared to individuals with less than five O’levels. Across the statistically significant m0odels the substantive interpretation is similar – with only a 1 per cent difference. With respect to sex, all models find sex to be statistically significant and hold similar substantive interpretation – with only a 1 per cent average marginal effects difference. With respect to housing tenure, only two models find housing tenure to be statistically significant – the model that recodes all educational attainment missingness equal to one and the imputation model. Of those models, both find housing tenure to have a very small impact upon an individual’s sorting into economic activity – around 1 per cent when translated to average marginal effects. Finally with respect to RGSC, whilst RGSC 1 is found to be statistically significant for the complete records analysis model – it is not statistically significant across any other models. All four models do find RGSC 3M to be statistically significant. Whilst the complete records analysis finds that individuals with an RGSC 3M social origins are 5 per cent increased probability of being in unemployment & OLF over education compared to individuals in RGSC 2 social origins, for all other models this substantive impact is instead 3 per cent. Finally, whilst the complete records analysis does not find any other RGSC categories to be statistically significant, all other models do find RGSC 5 to be statistically significant. All three models find there to be a 4-5 per cent increased probability of individuals from RGSC 5 social origins to enter into unemployment & OLF over education compared to their RGSC 2 social origins peers. The unemployment & OLF outcome category hosts the most amount of missingness – and the largest standard errors within the complete records analysis model. The relative deviation of all other models and their respective similarities with one another – with the exception of educational attainment[[25]](#footnote-25). – suggests that the complete records analysis model is not the best model fit for substantive interpretation of effects that structural inequalities may have on influencing individuals sorting into economic activity post-mandatory schooling.

Table 3.17 Comparison of Missingness across four models

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CRA RGSC  (Model 1) | | | Average Marginal Effects | | All Educational Attainment Missingness=0  (Model 2) | | | Average Marginal Effects | | All Educational Attainment Missingness=1  (Model 3) | | | Average Marginal Effects | | Imputed RGSC  (Model 4) | | | Average Marginal Effects | |
| Economic Activity | 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 |
| Employment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Less than five O’levels | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| Five or More O’levels | -1.38 | (0.24) | \*\*\* | -0.12 | (0.03) | -1.91 | (0.22) | \*\*\* | -0.17 | (0.02) | 0.11 | (0.16) |  | 0.01 | (0.03) | -1.34 | (0.23) | \*\*\* | -0.10 | (0.04) |
| Sex |  |  | \*\*\* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Female | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| Male | 0.30 | (0.22) |  | 0.02 | (0.03) | 0.61 | (0.15) | \*\*\* | 0.07 | (0.02) | 0.54 | (0.14) | \*\*\* | 0.07 | (0.02) | 0.64 | (0.14) | \*\*\* | 0.07 | (0.02) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Own Home | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| Don't Own Home | 0.30 | (0.26) |  | 0.03 | (0.03) | 0.54 | (0.16) | \*\* | 0.04 | (0.03) | 0.78 | (0.16) | \*\*\* | 0.07 | (0.03) | 0.58 | (0.16) | \*\*\* | 0.06 | (0.03) |
| RGSC |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 | -0.20 | (0.58) |  | -0.02 | (0.06) | -0.19 | (0.44) |  | 0.01 | (0.07) | -0.29 | (0.42) |  | -0.01 | (0.06) | -0.17 | (0.43) |  | 0.01 | (0.07) |
| 2 | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | Ref. | (.) |  | (.) | (.) |
| 3NM | -0.19 | (0.41) |  | -0.03 | (0.04) | 0.02 | (0.29) |  | -0.00 | (0.04) | 0.11 | (0.27) |  | 0.01 | (0.04) | 0.08 | (0.29) |  | 0.01 | (0.04) |
| 3M | 0.67 | (0.27) | \* | 0.06 | (0.04) | 0.81 | (0.20) | \*\*\* | 0.07 | (0.03) | 0.97 | (0.19) | \*\*\* | 0.10 | (0.03) | 0.88 | (0.20) | \*\*\* | 0.08 | (0.03) |
| 4 | 0.68 | (0.35) | \* | 0.07 | (0.05) | 0.54 | (0.24) | \* | 0.05 | (0.04) | 0.81 | (0.23) | \*\*\* | 0.09 | (0.04) | 0.62 | (0.25) | \* | 0.07 | (0.04) |
| 5 | 0.40 | (0.54) |  | 0.02 | (0.07) | 0.40 | (0.34) |  | -0.01 | (0.05) | 0.64 | (0.33) |  | 0.02 | (0.05) | 0.49 | (0.35) |  | 0.00 | (0.05) |
| Intercept | -1.32 | (0.25) | \*\*\* |  |  | -1.05 | (0.18) | \*\*\* |  |  | -1.71 | (0.20) | \*\*\* |  |  | -1.06 | (0.19) | \*\*\* |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Training & Apprenticeships |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Less than five O’levels | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| Five or More O’levels | -1.85 | (0.28) | \*\*\* | -0.17 | (0.03) | -2.42 | (0.26) | \*\*\* | -0.24 | (0.02) | 0.14 | (0.15) |  | 0.02 | (0.03) | -2.03 | (0.26) | \*\*\* | -0.22 | (0.03) |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Female | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| Male | 0.27 | (0.22) |  | 0.02 | (0.03) | 0.43 | (0.15) | \*\* | 0.02 | (0.02) | 0.35 | (0.14) | \* | 0.02 | (0.02) | 0.52 | (0.14) | \*\*\* | 0.03 | (0.02) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Own Home | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| Don't Own Home | 0.35 | (0.26) |  | 0.03 | (0.03) | 0.59 | (0.16) | \*\*\* | 0.06 | (0.03) | 0.86 | (0.15) | \*\*\* | 0.10 | (0.03) | 0.54 | (0.16) | \*\* | 0.04 | (0.03) |
| RGSC |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 | -0.67 | (0.78) |  | -0.06 | (0.05) | -1.17 | (0.64) |  | -0.14 | (0.05) | -1.29 | (0.62) | \* | -0.13 | (0.04) | -1.13 | (0.64) | \*\* | -0.13 | (0.05) |
| 2 | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | Ref. | (.) |  | (.) | (.) |
| 3NM | 0.44 | (0.38) |  | 0.06 | (0.05) | 0.16 | (0.28) |  | 0.03 | (0.04) | 0.26 | (0.26) |  | 0.04 | (0.04) | 0.19 | (0.28) |  | 0.03 | (0.04) |
| 3M | 0.75 | (0.29) | \* | 0.06 | (0.03) | 0.82 | (0.20) | \*\*\* | 0.07 | (0.03) | 1.00 | (0.19) | \*\*\* | 0.11 | (0.03) | 0.84 | (0.21) | \*\*\* | 0.07 | (0.03) |
| 4 | 0.76 | (0.36) | \* | 0.07 | (0.04) | 0.50 | (0.24) | \* | 0.04 | (0.04) | 0.79 | (0.23) | \*\*\* | 0.09 | (0.04) | 0.54 | (0.25) | \* | 0.05 | (0.04) |
| 5 | 0.80 | (0.52) |  | 0.08 | (0.07) | 0.83 | (0.32) | \*\* | 0.10 | (0.05) | 1.10 | (0.30) | \*\*\* | 0.15 | (0.05) | 0.90 | (0.35) | \*\* | 0.11 | (0.06) |
| Intercept | -1.39 | (0.26) | \*\*\* |  |  | -0.89 | (0.18) | \*\*\* |  |  | -1.63 | (0.20) | \*\*\* |  |  | -0.79 | (0.19) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Unemployment & Out of Labour Force |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Less than five O’levels | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| Five or More O’levels | -3.18 | (1.04) | \*\* | -0.04 | (0.01) | -3.50 | (1.02) | \*\*\* | -0.05 | (0.01) | -0.22 | (0.30) |  | -0.01 | (0.01) | -3.32 | (1.08) | \*\* | -0.06 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Female | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| Male | 1.02 | (0.48) | \* | 0.02 | (0.01) | 0.80 | (0.29) | \*\* | 0.02 | (0.01) | 0.74 | (0.29) | \*\* | 0.02 | (0.01) | 0.69 | (0.26) | \*\* | 0.01 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Own Home | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| Don't Own Home | 0.44 | (0.54) |  | 0.01 | (0.01) | 0.55 | (0.31) |  | 0.01 | (0.01) | 0.88 | (0.30) | \*\* | 0.02 | (0.01) | 0.68 | (0.29) | \* | 0.01 | (0.01) |
| RGSC |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 | 2.57 | (1.27) | \* | 0.07 | (0.05) | 0.59 | (0.84) |  | 0.03 | (0.04) | 0.46 | (0.82) |  | 0.02 | (0.03) | 0.64 | (0.86) |  | 0.04 | (0.05) |
| 2 | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| 3NM | -12.40 | (737.60) |  | -0.01 | (0.01) | -1.18 | (1.09) |  | -0.02 | (0.01) | -1.09 | (1.08) |  | -0.02 | (0.01) | -1.21 | (1.09) |  | -0.02 | (0.01) |
| 3M | 2.63 | (1.05) | \* | 0.05 | (0.01) | 1.24 | (0.44) | \*\* | 0.03 | (0.01) | 1.45 | (0.44) | \*\*\* | 0.03 | (0.01) | 1.32 | (0.46) | \*\* | 0.03 | (0.01) |
| 4 | 1.44 | (1.25) |  | 0.01 | (0.01) | 0.55 | (0.55) |  | 0.01 | (0.02) | 0.86 | (0.54) |  | 0.01 | (0.02) | 0.59 | (0.56) |  | 0.01 | (0.02) |
| 5 | 1.80 | (1.47) |  | 0.02 | (0.03) | 1.42 | (0.59) | \* | 0.04 | (0.03) | 1.73 | (0.58) | \*\* | 0.05 |  | 1.50 | (0.62) | \* | 0.05 | (0.03) |
| Intercept | -4.79 | (1.06) | \*\*\* |  |  | -3.16 | (0.43) | \*\*\* |  |  | -3.75 | (0.46) | \*\*\* |  |  | -2.88 | (0.43) | \*\*\* | (.) | (.) |
| Number of observations | 723 | | | | | 1368 | | | | | 1368 | | | | | 1645 | | | | |
| McFadden’s | 0.12 | | | | | 0.12 | | | | | 0.06 | | | | | - | | | | |
| McFadden’s Pseudo | 0.03 | | | | | 0.09 | | | | | 0.03 | | | | | - | | | | |
| Cox-Snell Pseudo | 0.20 | | | | | 0.25 | | | | | 0.13 | | | | | - | | | | |
| Nagelkerke Pseudo | 0.24 | | | | | 0.27 | | | | | 0.14 | | | | | - | | | | |
| AIC | 1319.08 | | | | | 2921.47 | | | | | 3121.11 | | | | | - | | | | |
| BIC | 1470.33 | | | | | 3062.44 | | | | | 3262.08 | | | | | - | | | | |
| Average RVI | - | | | | | - | | | | | - | | | | | 0.33 | | | | |
| Largest FMI | - | | | | | - | | | | | - | | | | | 0.62 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: BCS [Birth-Age 30]  Note: Comparison of Missingness across four models | | | | | | | | | | | | | | | | | | | | |

### Discussion and Conclusions

The fact that there are substantive differences between the complete records analysis and the imputed model suggests evidence for a MAR assumption. This indicates that missingness in these categories has a substantive impact on the resulting interpretation and analysis of results for the complete records analysis. The level of missingness present – almost entirely concentrated within the educational attainment variable within the model – suggest that data is missing at random, which has a direct impact on the ability to interpret results appropriately. The level of missingness that is directly responsible from a single variable – educational attainment – prompted the production and comparison of two other models alongside the complete records analysis and imputed model in an attempt to understand a rough interpretation of individuals sorting into economic activity post-mandatory schooling for the BCS. These two models focused on recoding all missingness for educational attainment at either zero or one (depending on the model) and comparing their substantive results to that of the imputation model to see if there were any distinct similarities or differences that may indicate the most accurate picture of true values. Of course, the ‘true’ values will never be known, they are missing. However, by using these two models as a point of comparison with the imputation model there now exists two polar extreme cases for the model, alongside the imputation model itself – which provides a rough outline of the possible substantive effects of individuals entering economic activity for the first time since the end of their mandatory schooling period.

When comparing all models together, it seems that the model that has recoded all missingness at educational attainment to zero – the most extreme model that states all missingness equals individuals receiving less than five O’levels, presents the most comparable substantive impacts when compared to the imputation model. This suggests that the level of missingness at educational attainment can be predicated upon individuals with less than five O’levels – in other words, individuals with worse educational outcomes – refusing or not answering the required survey instrument. This level of missingness has when looking at a direct comparison between complete records analysis and imputation models, resulted in a differing substantive interpretation between each model. If no handling of missing data were conducted, then the complete records analysis would be presented at face value. This implication would have meant erroneous conclusions would have been drawn from the complete records analysis, which simply do not exist with the imputed model. For example, a major inference from the complete records model is that the structural impacts on individual choice and opportunity have severely diminished within the BCS cohort. This when looking at the imputed model is simply incorrect. The substantive impact that sex, housing tenure, and social class has upon individual choice and opportunity is for some, very strong. Not only does this have implications for model interpretation and understanding the true nature of the role of structural influences upon choice and opportunity for BCS youth sorting into economic activity post-mandatory schooling. It also has wider implications for the great need of handling missing data practices to be widely implemented across social scientific research. Without using tools such as multiple imputation, results could be represented as novel – breaking the normal science barrier – when it could be a case of missing data impacting the substantive interpretation of our models – just as it has done with the BCS model.

## Discussion and Conclusions for Part 2

The overarching story of the British Cohort Study has been one of missing data. The attempt to duplicate the analysis from the previous part within this thesis has been disadvantaged by the level of missingness found across the proposed model. This from investigation appears to stem from two primary sources – the first and foremost issue comes from a lack of administrative foresight with handling the BCS cohort post-schooling. The tracking and tracing methods from the sweep at age 16 to the next sweep at age 26 resulted in a large number of lost contacts – many of whom are still lost. Secondly, the level of item missingness at educational attainment – the only variable from this model taken from a post-16 sweep severely limited the number of observations within the model itself – presenting large standard errors amongst the original complete records analysis.

At first, the resulting picture from the original complete records analysis seemed to suggest a rather egalitarian society that had emerged post-NCDS cohort. Where once there was large impactful influences of sex, social class, and housing tenure, the complete records analysis of the BCS youth suggested that there were no such influences upon individual choice and opportunity. Whilst this does in some respects support certain theories as to the death of social class (XXXX) as well as theories that argued the decline in influence of traditional structural inequalities (XXXX), it also presented issues for those theories as well. The lack of any influence of housing tenure within the complete records analysis stands in stark contrast to theories of new structuralism (XXXX). In fact, the results from the complete records analysis are suggestive of a novel finding within the field of youth transitions research. If this is where this analysis stopped, this would inevitably be the conclusion for this part of this thesis. However, this part initially dealt with a sensitivity analysis of social stratification measures of social origin. This analysis found that the initially proposed measure of NS-SEC was the least well-fitting measure of social stratification social origin within the model, instead preferring either CAMSIS or RGSC as alternative measures. Similar to the NCDS models presented in part one, the part two model using the CAMSIS model did not present any substantive impact of social stratification on individual sorting into economic activity. It was also the model that was most favoured by goodness-of-fit statistics. The overall bias that these statistics have for metric over categorical measures, as well as the overall marginal difference in said statistics between the CAMSIS and RGSC model meant that ultimately the RGSC model was chosen for substantive interpretation going forward. Whilst there was a small influence on social class upon sorting into economic activity, similar to the NS-SEC model, structural influences had a negligible if any impact upon individual choice and opportunity.

Lastly, part two of this thesis sought to deal with missing data within the BCS model. As previously stated, if this had not been attempted, the concluding remarks for this section would be entirely different. The resulting level of missingness within the BCS model demonstrated a need for imputation alongside other models as a point of comparison. These combined models seemed to confirm a MAR assumption. This suggests that the prior substantive interpretation of the complete records analysis may be erroneous. The interpretation of the imputed model are suggestive of a strong and lasting impact of structural influences upon choice and opportunity. This stands in direct contrast to theories that proclaim the ‘death of social class’ (XXXX) as well as new structuralist theories that suggest traditional structural inequalities are weakening and being replaced by newer structural cleavages such as housing tenure (XXXX). The resounding conclusion from the imputation model is that structural inequalities matter for influencing individuals’ choice and opportunity with respect to economic activity sorting post-mandatory schooling. Though the level of influence these structures have on individuals differs depending on the type of economic activity individuals sort into. The next part of this thesis seeks to explore the extent of these structural influences on yet another cohort from the 1990s onwards.

# The United Kingdom Household Panel Survey

## Introduction to Part 3

## Literature Review: UKHLS Timeframe and Context

### Story of transitions for UKHLS Youth

### Structural barriers to successful transitions – the role of social class and sex

## Data and Methods

### Introduction to the UKHLS data

### Synthetic Cohorts

### Introduction to measures for subsequent analysis

## Descriptive Statistics

## Modelling Main Economic Activity

### Discussion and Conclusions

## Sensitivity Analysis of Independent Variables

### Testing Measures of Parental Social Class

### Discussion and Conclusions

## Missing Data in the UKHLS

### Discussion and Conclusions

## Duplication Analysis using Next Steps

### Data and Methods

#### Introduction to Next Steps Data

#### Introduction to measures for subsequent analysis

### Descriptive Statistics

### Modelling Main Economic Activity

#### Discussion and Conclusions

### Sensitivity Analysis

#### Discussion and Conclusions

### Missing Data in Next Steps

#### Discussion and Conclusions

## Discussion and Conclusions for Part 3

# Comparison of NCDS, BCS, UKHLS, and Next Steps Cohorts

## Introduction to Part 4

## The effects of structural inequality across cohorts

### Discussion and Conclusions

# Conclusions

## Introduction to Part 5

## Substantive Conclusions

## Methodological Reflections

## Final Remarks

# Appendix:

## Appendix One: NCDS

Appendix 6.1.1 Goodness-of-fit summaries for explanatory variables and Economic Activity (CAMSIS)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 23040.88 | - | - | - | 23048.88 | 23077.03 |
| Null Model + Educational Attainment | 18829.53 | 4211.35 | 4 | 0.18 | 18845.83 | 18901.83 |
| Null Model + Sex | 22052.16 | 988.72 | 4 | 0.04 | 22068.16 | 22124.46 |
| Null Model + Tenure | 22224.78 | 816.11 | 4 | 0.04 | 22240.77 | 22297.07 |
| Null Model + CAMSIS | 21743.52 | 1297.36 | 8 | 0.06 | 21759.52 | 21815.82 |

Appendix 6.1.2 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (CAMSIS)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d.f. (from Previous) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 23040.88 | - | - | - | 23048.88 | 23077.03 |
| Null Model + Educational Attainment | 18829.53 | 4211.35 | 4 | 0.18 | 18845.83 | 18901.83 |
| Null Model + Educational Attainment + Sex | 17843.77 | 985.76 | 4 | 0.23 | 17867.77 | 17952.22 |
| Null Model + Educational Attainment + Sex + Tenure | 17602.92 | 240.85 | 4 | 0.24 | 17634.92 | 17747.51 |
| Null Model + Educational Attainment + Sex + Tenure + CAMSIS | 17374.46 | 228.46 | 20 | 0.25 | 17414.46 | 17555.21 |

Appendix 6.1.3 Goodness-of-fit summaries for explanatory variables and Economic Activity (RGSC)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d.f. (from Null) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 23150.59 | - | - | - | 23158.58 | 23186.75 |
| Null Model + Educational Attainment | 18908.96 | 4241.63 | 8 | 0.18 | 18924.96 | 18981.29 |
| Null Model + Sex | 22159.83 | 990.76 | 8 | 0.04 | 22175.83 | 22232.16 |
| Null Model + Tenure | 22328.74 | 821.85 | 8 | 0.03 | 22344.74 | 22401.07 |
| Null Model + RGSC | 21912.53 | 1128.35 | 24 | 0.05 | 21960.53 | 22129.42 |

Appendix 6.1.4 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (RGSC)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d.f. (from Previous) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 23040.88 | - | - | - | 23048.88 | 23077.03 |
| Null Model + Educational Attainment | 18829.53 | 4211.35 | 4 | 0.18 | 18845.83 | 18901.83 |
| Null Model + Educational Attainment + Sex | 17843.77 | 985.76 | 4 | 0.23 | 17867.77 | 17952.22 |
| Null Model + Educational Attainment + Sex + Tenure | 17602.92 | 240.85 | 4 | 0.24 | 17634.92 | 17747.51 |
| Null Model + Educational Attainment + Sex + Tenure | 17382.71 | 220.21 | 20 | 0.25 | 17454.71 | 17708.05 |

Appendix 6.1.5 Goodness-of-fit summaries for explanatory variables and Economic Activity (CAMSIS) using SOC2000 and SOC90 Codes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d.f. (from Null) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 23150.59 | - | - | - | 23158.58 | 23186.75 |
| Null Model + Educational Attainment | 18908.96 | 4241.63 | 8 | 0.18 | 18924.96 | 18981.29 |
| Null Model + Sex | 22159.83 | 990.76 | 8 | 0.04 | 22175.83 | 22232.16 |
| Null Model + Tenure | 22328.74 | 821.85 | 8 | 0.03 | 22344.74 | 22401.07 |
| Null Model + CAMSIS (SOC 2000) | 21743.52 | 1297.36 | 8 | 0.06 | 21759.52 | 21815.82 |
| Null Model + CAMSIS (SOC 90) | 22199.41 | 841.47 | 8 | 0.04 | 22215.41 | 22271.71 |

Appendix 6.1.6 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (CAMSIS) using SOC2000 and SOC90 Codes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 23040.88 | - | - | - | 23048.88 | 23077.03 |
| Null Model + Educational Attainment | 18829.53 | 4211.35 | 4 | 0.18 | 18845.83 | 18901.83 |
| Null Model + Educational Attainment + Sex | 17843.77 | 985.76 | 4 | 0.23 | 17867.77 | 17952.22 |
| Null Model + Educational Attainment + Sex + Tenure | 17602.92 | 240.85 | 4 | 0.24 | 17634.92 | 17747.51 |
| Null Model + Educational Attainment + Sex + Tenure + CAMSIS (SOC 2000) | 17374.46 | 228.46 | 20 | 0.25 | 17414.46 | 17555.21 |
| Null Model + Educational Attainment + Sex + Tenure + CAMSIS (SOC 90) | 17471.40 | 131.52 | 20 | 0.24 | 17511.40 | 17652.15 |

Appendix 6.1.7 Goodness-of-fit summaries for explanatory variables and Economic Activity (RGSC) using SOC2000 and SOC90 Codes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d.f. (from Null) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 23150.59 | - | - | - | 23158.58 | 23186.75 |
| Null Model + Educational Attainment | 18908.96 | 4241.63 | 8 | 0.18 | 18924.96 | 18981.29 |
| Null Model + Sex | 22159.83 | 990.76 | 8 | 0.04 | 22175.83 | 22232.16 |
| Null Model + Tenure | 22328.74 | 821.85 | 8 | 0.03 | 22344.74 | 22401.07 |
| Null Model + RGSC (SOC 2000) | 21912.53 | 1128.35 | 24 | 0.05 | 21960.53 | 22129.42 |
| Null Model + RGSC (SOC 90) |  |  | 24 |  |  |  |

Appendix 6.1.8 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (RGSC) using SOC200 and SOC90 Codes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 23040.88 | - | - | - | 23048.88 | 23077.03 |
| Null Model + Educational Attainment | 18829.53 | 4211.35 | 4 | 0.18 | 18845.83 | 18901.83 |
| Null Model + Educational Attainment + Sex | 17843.77 | 985.76 | 4 | 0.23 | 17867.77 | 17952.22 |
| Null Model + Educational Attainment + Sex + Tenure | 17602.92 | 240.85 | 4 | 0.24 | 17634.92 | 17747.51 |
| Null Model + Educational Attainment + Sex + Tenure + RGSC (SOC 2000) | 17382.71 | 220.21 | 20 | 0.25 | 17454.71 | 17708.05 |
| Null Model + Educational Attainment + Sex + Tenure + RGSC (SOC 90) | 17409.54 | 193.38 | 20 | 0.24 | 17481.54 | 17734.88 |

Appendix 6.1.9 Sensitivity Analysis of SOC Codes (CAMSIS)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SOC 2000 Codes | | | | | SOC 90 Codes | | | | |
|  | CAMSIS | | | Average Marginal Effects | | CAMSIS | | | Average Marginal Effects | |
| Economic Activity: ‘School’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** |
| Employment |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.57 | (0.08) | \*\*\* | -0.38 | (0.01) | -3.63 | (0.08) | \*\*\* | -0.40 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* |  |  |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | -0.39 | (0.07) | \*\*\* | -0.17 | (0.01) | -0.41 | (0.07) | \*\*\* | -0.17 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* |  |  |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.65 | (0.08) | \*\*\* | 0.08 | (0.01) | 0.79 | (0.08) | \*\*\* | 0.09 | (0.01) |
| CAMSIS | -0.04 | (0.00) | \*\*\* | -0.00 | (0.00) | -0.03 | (0.00) | \*\*\* | -0.00 | (0.00) |
| Intercept | 3.45 | (0.16) | \*\*\* | (.) | (.) | 2.89 | (0.15) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Non-Traditional Education |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -1.70 | (0.09) | \*\*\* | 0.02 | (0.01) | -1.71 | (0.09) | \*\*\* | 0.03 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | -0.83 | (0.09) | \*\*\* | -0.07 | (0.01) | -0.84 | (0.09) | \*\*\* | -0.07 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | -0.12 | (0.10) |  | -0.04 | (0.01) | -0.09 | (0.09) |  | -0.04 | (0.01) |
| CAMSIS | -0.01 | (0.00) |  | 0.00 | (0.00) | -0.01 | (0.00) |  | 0.00 | (0.00) |
| Intercept | 0.84 | (0.19) | \*\*\* | (.) | (.) | 0.76 | (0.18) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| School | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Training & Apprenticeships |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.21 | (0.09) | \*\*\* | -0.17 | (0.01) | -3.26 | (0.09) | \*\*\* | -0.17 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 1.52 | (0.09) | \*\*\* | 0.24 | (0.01) | 1.51 | (0.09) | \*\*\* | 0.24 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.34 | (0.09) | \*\*\* | -0.01 | (0.01) | 0.45 | (0.08) | \*\*\* | -0.01 | (0.01) |
| CAMSIS | -0.04 | (0.00) | \*\*\* | 0.00 | (0.00) | -0.03 | (0.00) | \*\*\* | -0.00 | (0.00) |
| Intercept | 1.53 | (0.18) | \*\*\* | (.) | (.) | 1.08 | (0.17) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Unemployment & Out of Labour Force |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.93 | (0.27) | \*\*\* | -0.03 | (0.00) | -3.99 | (0.27) | \*\*\* | -0.03 | (0.00) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | -0.36 | (0.14) | \* | -0.01 | (0.00) | -0.38 | (0.14) | \*\* | -0.01 | (0.00) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.81 | (0.16) | \*\*\* | 0.01 | (0.00) | 0.96 | (0.16) | \*\*\* | 0.01 | (0.01) |
| CAMSIS | -0.05 | (0.01) | \*\*\* | -0.00 | (0.00) | -0.05 | (0.01) | \*\*\* | -0.00 | (0.00) |
| Intercept | 1.35 | (0.33) | \*\*\* | (.) | (.) | 0.87 | (0.32) | \*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Number of observations | 8411 | | | | | 8411 | | | | |
| McFadden’s | 0.25 | | | | | 0.24 | | | | |
| McFadden’s Adjusted Pseudo | 0.24 | | | | | 0.24 | | | | |
| Cox-Snell Pseudo | 0.49 | | | | | 0.48 | | | | |
| Nagelkerke Pseudo | 0.52 | | | | | 0.52 | | | | |
| AIC | 17414.46 | | | | | 17511.40 | | | | |
| BIC | 17555.21 | | | | | 17652.15 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: Complete Records Analysis | | | | | | | | | | |

Appendix 6.1.10 Sensitivity Analysis of SOC Codes (RGSC)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SOC 2000 Codes | | | | | SOC 90 Codes | | | | |
|  | RGSC | | | Average Marginal Effects | | RGSC | | | Average Marginal Effects | |
| Economic Activity: ‘School’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** |
| Employment |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.60 | (0.08) | \*\*\* | -0.39 | (0.01) | -3.62 | (0.08) | \*\*\* | -0.39 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* |  |  |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | -0.40 | (0.07) | \*\*\* | -0.17 | (0.01) | -0.41 | (0.07) | \*\*\* | -0.17 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* |  |  |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.71 | (0.08) | \*\*\* | 0.08 | (0.01) | 0.78 | (0.08) | \*\*\* | 0.08 | (0.01) |
| RGSC |  |  |  |  |  |  |  |  |  |  |
| *1* | -0.97 | (0.23) | \*\*\* | -0.10 | (0.03) | -0.32 | (0.26) |  | -0.10 | (0.03) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3NM* | -0.01 | (0.13) |  | -0.01 | (0.02) | 0.57 | (0.17) | \*\*\* | -0.01 | (0.02) |
| *4M* | 0.70 | (0.10) | \*\*\* | 0.05 | (0.01) | 1.29 | 1.29 | \*\*\* | 0.05 | (0.01) |
| *4* | 0.79 | (0.13) | \*\*\* | 0.06 | (0.02) | 1.02 | (0.15) | \*\*\* | 0.06 | (0.02) |
| *5* | 1.14 | (0.16) | \*\*\* | 0.12 | (0.02) | 1.18 | (0.19) | \*\*\* | 0.12 | (0.02) |
| Intercept | 1.17 | (0.10) | \*\*\* | (.) | (.) | 0.66 | (0.15) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Non-Traditional Education |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -1.71 | (0.09) | \*\*\* | 0.03 | (0.01) | -1.73 | (0.09) | \*\*\* | 0.03 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | -0.84 | (0.09) | \*\*\* | -0.07 | (0.01) | -0.84 | (0.09) | \*\*\* | -0.07 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | -0.08 | (0.10) |  | -0.04 | (0.01) | -0.03 | (0.09) |  | -0.04 | (0.01) |
| RGSC |  |  |  |  |  |  |  |  |  |  |
| *1* | -0.49 | (0.21) | \* | 0.00 | (0.02) | -0.31 | (0.24) |  | 0.00 | (0.02) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3NM* | 0.12 | (0.14) |  | 0.01 | (0.01) | 0.28 | (0.16) |  | 0.01 | (0.01) |
| *4M* | 0.23 | (0.11) | \* | -0.02 | (0.01) | 0.17 | (0.16) |  | -0.02 | (0.01) |
| *4* | 0.11 | (0.16) |  | -0.03 | (0.01) | 0.27 | (0.15) |  | -0.03 | (0.01) |
| *5* | 0.10 | (0.23) |  | -0.04 | (0.01) | 0.04 | (0.22) |  | -0.04 | (0.01) |
| Intercept | 0.18 | (0.11) |  | (.) | (.) | 0.09 | (0.15) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| School | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Training & Apprenticeships |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.25 | (0.09) | \*\*\* | -0.17 | (0.01) | -3.27 | (0.09) | \*\*\* | -0.17 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 1.52 | (0.09) | \*\*\* | 0.24 | (0.01) | 1.51 | (0.09) | \*\*\* | 0.24 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.40 | (0.08) | \*\*\* | -0.01 | (0.01) | 0.47 | (0.08) | \*\*\* | -0.01 | (0.01) |
| RGSC |  |  |  |  |  |  |  |  |  |  |
| *1* | -0.52 | (0.23) | \* | 0.01 | (0.03) | -0.33 | (0.28) |  | 0.01 | (0.03) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3NM* | 0.14 | (0.15) |  | 0.02 | (0.02) | 0.44 | (0.18) | \* | 0.02 | (0.02) |
| *4M* | 0.78 | (0.11) | \*\*\* | 0.04 | (0.01) | 1.04 | (0.17) | \*\*\* | 0.04 | (0.01) |
| *4* | 0.83 | (0.14) | \*\*\* | 0.04 | (0.01) | 0.87 | (0.17) | \*\*\* | 0.04 | (0.01) |
| *5* | 0.76 | (0.19) | \*\*\* | -0.00 | (0.02) | 0.76 | (0.21) | \*\*\* | -0.00 | (0.02) |
| Intercept | -0.56 | (0.12) | \*\*\* | (.) | (.) | -0.84 | (0.17) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Unemployment & Out of Labour Force |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.97 | (0.27) | \*\*\* | -0.03 | (0.00) | -4.01 | (0.27) | \*\*\* | -0.03 | (0.00) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | -0.37 | (0.14) | \* | -0.01 | (0.00) | -0.38 | (0.14) | \*\* | -0.01 | (0.00) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.91 | (0.16) | \*\*\* | 0.01 | (0.00) | 0.99 | (0.16) | \*\*\* | 0.01 | (0.01) |
| RGSC |  |  |  |  |  |  |  |  |  |  |
| *1* | -2.02 | (1.03) | \* | -0.02 | (0.01) | -0.61 | (0.67) |  | -0.02 | (0.01) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3NM* | -0.91 | (0.39) | \* | -0.02 | (0.01) | -0.87 | (0.47) |  | -0.02 | (0.01) |
| *4M* | 0.48 | (0.22) | \* | -0.00 | (0.01) | 0.68 | (0.33) | \* | -0.00 | (0.01) |
| *4* | 0.91 | (0.25) | \*\*\* | 0.01 | (0.01) | 0.79 | (0.33) | \* | 0.01 | (0.01) |
| *5* | 1.19 | (0.28) | \*\*\* | 0.01 | (0.01) | 0.96 | 0.96 | \*\* | 0.01 | (0.01) |
| Intercept | 1.35 | (0.21) | \*\*\* | (.) | (.) | -1.57 | (0.32) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Number of observations | 8411 | | | | | 8411 | | | | |
| McFadden’s | 0.25 | | | | | 0.24 | | | | |
| McFadden’s Adjusted Pseudo | 0.24 | | | | | 0.24 | | | | |
| Cox-Snell Pseudo | 0.49 | | | | | 0.49 | | | | |
| Nagelkerke Pseudo | 0.52 | | | | | 0.52 | | | | |
| AIC | 17454.71 | | | | | 17481.54 | | | | |
| BIC | 17708.05 | | | | | 17734.88 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: Complete Records Analysis | | | | | | | | | | |

## Appendix Two: BCS

Appendix 6.2.1 Goodness-of-fit summaries for explanatory variables and Economic Activity (CAMSIS)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 1418.18 | - | - | - | 1424.18 | 1437.93 |
| Null Model + Educational Attainment | 1305.16 | 113.02 | 3 | 0.08 | 1317.16 | 1344.66 |
| Null Model + Sex | 1414.68 | 3.5 | 3 | 0.00 | 1426.68 | 1454.18 |
| Null Model + Tenure | 1403.27 | 14.91 | 3 | 0.01 | 1415.27 | 1442.77 |
| Null Model + CAMSIS | 1377.94 | 40.24 | 3 | 0.03 | 1389.94 | 1417.44 |

Appendix 6.2.2 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (CAMSIS)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d.f. (from Previous) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 1418.18 | - | - | - | 1424.18 | 1437.93 |
| Null Model + Educational Attainment | 1305.16 | 113.02 | 3 | 0.08 | 1317.16 | 1344.66 |
| Null Model + Educational Attainment + Sex | 1299.29 | 5.87 | 3 | 0.08 | 1317.29 | 1358.54 |
| Null Model + Educational Attainment + Sex + Tenure | 1294.22 | 5.07 | 3 | 0.09 | 1318.22 | 1373.23 |
| Null Model + Educational Attainment + Sex + Tenure + CAMSIS | 1279.34 | 14.88 | 3 | 0.10 | 1309.34 | 1378.09 |

Appendix 6.2.3 Goodness-of-fit summaries for explanatory variables and Economic Activity (RGSC)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d.f. (from Null) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 1418.18 | - | - | - | 1424.18 | 1437.93 |
| Null Model + Educational Attainment | 1305.16 | 113.02 | 3 | 0.08 | 1317.16 | 1344.66 |
| Null Model + Sex | 1414.68 | 3.5 | 3 | 0.00 | 1426.68 | 1454.18 |
| Null Model + Tenure | 1403.27 | 14.91 | 3 | 0.01 | 1415.27 | 1442.77 |
| Null Model + RGSC | 1362.61 | 55.57 | 15 | 0.04 | 1398.61 | 1481.11 |

Appendix 6.2.4 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (RGSC)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d.f. (from Previous) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 1418.18 | - | - | - | 1424.18 | 1437.93 |
| Null Model + Educational Attainment | 1305.16 | 113.02 | 3 | 0.08 | 1317.16 | 1344.66 |
| Null Model + Educational Attainment + Sex | 1299.29 | 5.87 | 3 | 0.08 | 1317.29 | 1358.54 |
| Null Model + Educational Attainment + Sex + Tenure | 1294.22 | 5.07 | 3 | 0.09 | 1318.22 | 1373.23 |
| Null Model + Educational Attainment + Sex + Tenure + RGSC | 1258.82 | 34.40 | 15 | 0.11 | 1312.82 | 1436.57 |

Appendix 6.2.5 Goodness-of-fit summaries for explanatory variables and Economic Activity (CAMSIS) using SOC2000 and SOC90 Codes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d.f. (from Null) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 1418.18 | - | - | - | 1424.18 | 1437.93 |
| Null Model + Educational Attainment | 1305.16 | 113.02 | 3 | 0.08 | 1317.16 | 1344.66 |
| Null Model + Sex | 1414.68 | 3.5 | 3 | 0.00 | 1426.68 | 1454.18 |
| Null Model + Tenure | 1403.27 | 14.91 | 3 | 0.01 | 1415.27 | 1442.77 |
| Null Model + CAMSIS (SOC 2000) | 1373.37 | 44.82 | 3 | 0.03 | 1385.37 | 1412.87 |
| Null Model + CAMSIS (SOC 90) | 1377.94 | 40.24 | 3 | 0.03 | 1389.94 | 1417.44 |

Appendix 6.2.6 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (CAMSIS) using SOC2000 and SOC90 Codes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 1418.18 | - | - | - | 1424.18 | 1437.93 |
| Null Model + Educational Attainment | 1305.16 | 113.02 | 3 | 0.08 | 1317.16 | 1344.66 |
| Null Model + Educational Attainment + Sex | 1299.29 | 5.87 | 3 | 0.08 | 1317.29 | 1358.54 |
| Null Model + Educational Attainment + Sex + Tenure | 1294.22 | 5.07 | 3 | 0.09 | 1318.22 | 1373.23 |
| Null Model + Educational Attainment + Sex + Tenure + CAMSIS (SOC 2000) | 1278.04 | 16.18 | 3 | 0.10 | 1308.04 | 1376.79 |
| Null Model + Educational Attainment + Sex + Tenure + CAMSIS (SOC 90) | 1279.34 | 14.88 | 3 | 0.10 | 1309.34 | 1378.09 |

Appendix 6.2.7 Goodness-of-fit summaries for explanatory variables and Economic Activity (NS-SEC) using SOC2000 and SOC90 Codes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d.f. (from Null) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 1418.18 | - | - | - | 1424.18 | 1437.93 |
| Null Model + Educational Attainment | 1305.16 | 113.02 | 3 | 0.08 | 1317.16 | 1344.66 |
| Null Model + Sex | 1414.68 | 3.5 | 3 | 0.00 | 1426.68 | 1454.18 |
| Null Model + Tenure | 1403.27 | 14.91 | 3 | 0.01 | 1415.27 | 1442.77 |
| Null Model + NS-SEC (SOC 2000) | 1353.42 | 64.76 | 21 | 0.05 | 1401.42 | 1511.42 |
| Null Model + NS-SEC (SOC 90) | 1353.93 | 64.26 | 21 | 0.05 | 1401.93 | 1511.93 |

Appendix 6.2.8 Model building goodness-of-fit summaries for multiple logistic regression model of Economic Activity (NS-SEC) using SOC2000 and SOC90 Codes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Adjusted Pseudo | AIC | BIC |
| Null Model | 1418.18 | - | - | - | 1424.18 | 1437.93 |
| Null Model + Educational Attainment | 1305.16 | 113.02 | 3 | 0.08 | 1317.16 | 1344.66 |
| Null Model + Educational Attainment + Sex | 1299.29 | 5.87 | 3 | 0.08 | 1317.29 | 1358.54 |
| Null Model + Educational Attainment + Sex + Tenure | 1294.22 | 5.07 | 3 | 0.09 | 1318.22 | 1373.23 |
| Null Model + Educational Attainment + Sex + Tenure + NS-SEC (SOC 2000) | 1253.08 | 41.14 | 21 | 0.12 | 1319.08 | 1470.33 |
| Null Model + Educational Attainment + Sex + Tenure + NS-SEC (SOC 90) | 1253.50 | 40.72 | 21 | 0.12 | 1319.50 | 1470.75 |

Appendix 6.2.9 Sensitivity Analysis of SOC Codes (CAMSIS)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SOC 2000 Codes | | | | | SOC 90 Codes | | | | |
|  | CAMSIS | | | Average Marginal Effects | | CAMSIS | | | Average Marginal Effects | |
| Economic Activity: ‘School’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** |
| Employment |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -1.33 | (0.24) | \*\*\* | -0.12 | (0.03) | -1.35 | (0.24) | \*\*\* | -0.12 | (0.03) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* |  |  |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 0.31 | (0.22) |  | 0.03 | (0.03) | 0.31 | (0.22) |  | 0.03 | (0.03) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* |  |  |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.25 | (0.25) |  | 0.02 | (0.03) | 0.27 | (0.25) |  | 0.02 | (0.03) |
| CAMSIS | -0.02 | (0.01) | \*\* | -0.00 | (0.00) | -0.02 | (0.01) | \*\* | -0.00 | (0.00) |
| Intercept | 0.15 | (0.43) |  | (.) | (.) | 0.05 | (0.41) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| School | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Training & Apprenticeships |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -1.80 | (0.28) | \*\*\* | -0.16 | (0.03) | -1.81 | (0.28) | \*\*\* | -0.16 | (0.03) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 0.25 | (0.22) |  | 0.01 | (0.03) | 0.26 | (0.22) |  | 0.02 | (0.03) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.36 | (0.25) |  | 0.04 | (0.03) | 0.36 | (0.25) |  | 0.04 | (0.03) |
| CAMSIS | -0.03 | (0.01) | \*\* | -0.00 | (0.00) | -0.03 | (0.01) | \*\* | -0.00 | (0.00) |
| Intercept | 0.34 | (0.45) |  | (.) | (.) | 0.30 | (0.43) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Unemployment & Out of Labour Force |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.04 | (1.04) | \*\* | -0.04 | (0.01) | -3.09 | (1.04) | \*\* | -0.04 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 0.98 | (0.48) | \* | 0.02 | (0.01) | 0.98 | (0.48) | \* | 0.02 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.18 | (0.52) |  | 0.00 | (0.01) | 0.23 | (0.52) |  | 0.00 | (0.01) |
| CAMSIS | -0.04 | (0.02) | \* | -0.00 | (0.00) | -0.03 | (0.02) |  | -0.00 | (0.00) |
| Intercept | -0.99 | (0.98) |  | (.) | (.) | -1.43 | (0.92) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Number of observations | 723 | | | | | 723 | | | | |
| McFadden’s | 0.10 | | | | | 0.10 | | | | |
| McFadden’s Adjusted Pseudo | 0.05 | | | | | 0.05 | | | | |
| Cox-Snell Pseudo | 0.18 | | | | | 0.18 | | | | |
| Nagelkerke Pseudo | 0.21 | | | | | 0.20 | | | | |
| AIC | 1308.04 | | | | | 1309.34 | | | | |
| BIC | 1376.79 | | | | | 1378.09 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: BCS [Birth-Age 30]  Note: Complete Records Analysis | | | | | | | | | | |

Appendix 6.2.10 Sensitivity Analysis of SOC Codes (NS-SEC)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SOC 2000 Codes | | | | | SOC 90 Codes | | | | |
|  | NS-SEC | | | Average Marginal Effects | | NS-SEC | | | Average Marginal Effects | |
| Economic Activity: ‘School’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** |
| Employment |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -1.37 | (0.24) | \*\*\* | -0.12 | (0.03) | -1.40 | (0.25) | \*\*\* | -0.12 | (0.03) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 0.33 | (0.22) |  | 0.03 | (0.03) | 0.32 | (0.22) |  | 0.03 | (0.03) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.25 | (0.26) |  | 0.02 | (0.03) | 0.26 | (0.26) |  | 0.02 | (0.03) |
| NS-SEC |  |  |  |  |  |  |  |  |  |  |
| *1.1* | -1.32 | (0.78) |  | -0.12 | (0.05) | -1.84 | (1.05) |  | -0.12 | (0.04) |
| *1.2* | -0.54 | (0.54) |  | -0.06 | (0.06) | 0.28 | (0.46) |  | 0.03 | (0.06) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3* | -0.23 | (0.43) |  | -0.04 | (0.05) | -0.08 | (0.43) |  | -0.02 | (0.05) |
| *4* | 0.31 | (0.42) |  | -0.01 | (0.05) | 0.60 | (0.42) |  | 0.03 | (0.05) |
| *5* | 0.55 | (0.35) |  | 0.03 | (0.05) | 0.76 | (0.35) |  | 0.06 | (0.05) |
| *6* | 0.35 | (0.38) |  | 0.02 | (0.05) | 0.50 | (0.39) |  | 0.03 | (0.05) |
| *7* | 0.46 | (0.35) |  | 0.04 | (0.05) | 0.54 | (0.37) |  | 0.05 | (0.05) |
| Intercept | -1.14 | (0.28) | \*\*\* | (.) | (.) | -1.30 | (0.28) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| School | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Training & Apprenticeships |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -1.87 | (0.28) | \*\*\* | -0.17 | (0.03) | -1.86 | (0.28) | \*\*\* | -0.16 | (0.03) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 0.27 | (0.23) |  | 0.01 | (0.03) | 0.26 | (0.23) |  | 0.01 | (0.03) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.36 | (0.26) |  | 0.04 | (0.03) | 0.33 | (0.26) |  | 0.03 | (0.03) |
| NS-SEC |  |  |  |  |  |  |  |  |  |  |
| *1.1* | 0.08 | (0.58) |  | 0.03 | (0.06) | 0.11 | (0.64) |  | 0.03 | (0.06) |
| *1.2* | -0.94 | (0.79) |  | -0.06 | (0.04) | 0.07 | (0.63) |  | -0.00 | (0.05) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3* | 0.41 | (0.45) |  | 0.05 | (0.05) | 0.65 | (0.47) |  | 0.07 | (0.05) |
| *4* | 1.09 | (0.44) | \* | 0.11 | (0.05) | 1.34 | (0.45) | \*\* | 0.13 | (0.05) |
| *5* | 1.08 | (0.39) | \*\* | 0.11 | (0.05) | 1.34 | (0.41) | \*\*\* | 0.13 | (0.04) |
| *6* | 0.72 | (0.43) |  | 0.07 | (0.05) | 1.06 | (0.43) | \* | 0.10 | (0.05) |
| *7* | 0.63 | (0.40) |  | 0.05 | (0.04) | 0.87 | (0.43) | \* | 0.08 | (0.04) |
| Intercept | -1.48 | (0.32) | \*\*\* | (.) | (.) | -1.71 | (0.35) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Unemployment & Out of Labour Force |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Five or More O’levels* | -3.18 | (1.04) | \*\* | -0.04 | (0.01) | -3.25 | (1.04) | \*\* | -0.04 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Male* | 1.03 | (0.48) | \* | 0.02 | (0.01) | 1.01 | (0.48) | \* | 0.02 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | 0.26 | (0.54) |  | 0.00 | (0.01) | 0.35 | (0.54) |  | 0.01 | (0.01) |
| NS-SEC |  |  |  |  |  |  |  |  |  |  |
| *1.1* | -13.07 | (1070.37) |  | -0.01 | (0.01) | -14.19 | (1165.50) |  | -0.03 | (0.01) |
| *1.2* | 1.64 | (1.26) |  | 0.05 | (0.04) | 0.81 | (0.96) |  | 0.03 | (0.04) |
| *2* | Ref. | (.) |  | (.) | (.) | (.) | (.) |  | (.) | (.) |
| *3* | -12.79 | (759.94) |  | -0.01 | (0.01) | -13.90 | (813.39) |  | -0.03 | (0.01) |
| *4* | 2.42 | (1.13) | \* | 0.06 | (0.03) | 1.14 | (0.82) |  | 0.02 | (0.03) |
| *5* | 1.71 | (1.14) |  | 0.02 | (0.02) | 0.94 | (0.77) |  | 0.01 | (0.02) |
| *6* | 1.60 | (1.16) |  | 0.02 | (0.02) | 0.59 | (0.82) |  | 0.01 | (0.02) |
| *7* | 1.35 | (1.15) |  | 0.02 | (0.02) | -0.26 | (0.95) |  | -0.01 | (0.02) |
| Intercept | -4.23 | (1.05) | \*\*\* | (.) | (.) | -3.25 | (0.67) | \*\*\* | (.) | (.) |
|  |  |  |  |  |  |  |  |  |  |  |
| Number of observations | 723 | | | | | 723 | | | | |
| McFadden’s | 0.12 | | | | | 0.12 | | | | |
| McFadden’s Adjusted Pseudo | 0.03 | | | | | 0.03 | | | | |
| Cox-Snell Pseudo | 0.20 | | | | | 0.20 | | | | |
| Nagelkerke Pseudo | 0.24 | | | | | 0.24 | | | | |
| AIC | 1319.08 | | | | | 1319.50 | | | | |
| BIC | 1470.33 | | | | | 1470.75 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: BCS [Birth-Age 30]  Note: Complete Records Analysis | | | | | | | | | | |

# 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 College London, UCL Institute of Education, Centre for Longitudinal Studies, 2023, *Next Steps: Sweeps 1-8, 2004-2016*, [data collection], UK Data Service, *16th Edition,*Accessed 22 February 2024. SN: 5545, [DOI: http://doi.org/10.5255/UKDA-SN-5545-8](http://doi.org/10.5255/UKDA-SN-5545-8)

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)

Reference List

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.

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.

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.

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.

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].

Brooks, R. (2009) *Transitions from education to work: new perspectives from Europe and beyond*. Springer.

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. 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].

Calderwood, L. and Sanchez, C. (2016) ‘Next Steps (formerly known as the Longitudinal Study of Young People in England)’, *Open Health Data*, 4, p. e2. Available at: https://doi.org/10.5334/ohd.16.

Carpenter, J.R. and Kenward, M. (2012) *Multiple imputation and its application*. John Wiley & Sons.

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.

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.

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, 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.

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., 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. (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, 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.

Firth, D. (2000) ‘QV Calculator : Quasi-variances in Xlisp-Stat and on the Web’, *Journal of Statistical Software* [Preprint].

Firth, D. (2003) ‘Overcoming the Reference Category Problem in the Presentation of Statistical Models’, *Sociological Methodology*, 33(1), pp. 1–18. Available at: https://doi.org/10.1111/j.0081-1750.2003.t01-1-00125.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. 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.

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. (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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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].

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.

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.

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. (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.

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.

ONS (2023) *Unemployment rate*. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment/timeseries/mgsx/lms (Accessed: 1 May 2023).

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).

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.

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.

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. (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.

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, T.J. and McKenna, C.M. (2013) ‘A Comparison of Logistic Regression Pseudo R2 Indices’, 39.

*SOC 2000 - Office for National Statistics* (2000). Available at: https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/socarchive (Accessed: 8 January 2024).

*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.

Stewart, A., Prandy, K. and Blackburn, R.M. (1973) ‘Measuring the Class Structure’, *Nature*, 245(5426), pp. 415–417. Available at: https://doi.org/10.1038/245415a0.

Stewart, A., Prandy, K. and Blackburn, R.M. (1980) *Social Stratification and Occupations*. Springer.

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.

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.

Treiman, D.J. (2009) *Quantitative data analysis doing social research to test ideas*. Jossey-Bass.

University College London (2022) ‘1970 British Cohort Study: Activity Histories 1986-2016’, *Centre for Longitudinal Studies* [Preprint]. Available at: https://doi.org/10.5255/UKDA-SN-6943-4.

University College London, UCL Institute of Education, and Centre for Longitudinal Studies (2023) ‘National Child Development Study.’, *UK Data Service* [Preprint], (13th Release). Available at: https://doi.org/DOI: http://doi.org/10.5255/UKDA-Series-2000032.

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.

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.

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. O’grades for Scotland [↑](#footnote-ref-1)
2. 13th Release [↑](#footnote-ref-2)
3. 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-3)
4. This latter category can be considered an ‘Other’ category. [↑](#footnote-ref-4)
5. ‘’Rule6’’ means N/A. [↑](#footnote-ref-5)
6. Either in Social Housing or privately rented accommodation. [↑](#footnote-ref-6)
7. Appendix 1, table 6.1.5 onwards [↑](#footnote-ref-7)
8. 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-8)
9. For these reasons when it comes to the sensitivity analysis, RGSC 2 will be used as the reference category of choice [↑](#footnote-ref-9)
10. https://warwick.ac.uk/fac/sci/statistics/staff/academic-research/firth/software/qvcalc/kuvee/ [↑](#footnote-ref-10)
11. This refers to the housing tenure meausre for the employment category – in SOC 2000 the average margianl effect is 8 per cent and in the SOC 90 construction it is 9 per cent [↑](#footnote-ref-11)
12. Discussed at length in section below on Multiple Imputation [↑](#footnote-ref-12)
13. The seed for MI model is 12346, it can also be found in the .do file within the Github page. [↑](#footnote-ref-13)
14. 50 imputations were used in MI models. [↑](#footnote-ref-14)
15. Variable n4118 used [↑](#footnote-ref-15)
16. The variable in question was acatnn236, a categorical variable. [↑](#footnote-ref-16)
17. Burn-in was 20 during imputation. [↑](#footnote-ref-17)
18. Former President of the European Commission speaking about the dramatic change of British society in the Guardian. [↑](#footnote-ref-18)
19. Other includes those respondents that cannot be accurately traced through any of the aforementioned categories. [↑](#footnote-ref-19)
20. Percentages are based on the participants divided by total cohort. [↑](#footnote-ref-20)
21. 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-21)
22. Age 26 was the first time the cohort member themselves were in complete control of answering the survey itself [↑](#footnote-ref-22)
23. Up to age 16, cohort members were traced through school records. After many left the school environment, it became very difficult to accurately track their location. A large effort was put in for the age 30 sweep of the BCS to regain some of these lost cohort members. [↑](#footnote-ref-23)
24. The rseed of the imputation is ‘12346’ [↑](#footnote-ref-24)
25. This is to be expected considering that missingness is being recoded as zero or one on the educational attainment variable for two of the models. [↑](#footnote-ref-25)