Essays in Poverty and Household Analysis in Africa

By

Rumman Khan

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy.

June 2017

Abstract

This thesis consists of three self-contained chapters on issues relating to the empirics of poverty and household analysis in Africa. The first analyses poverty at the country level, highlighting the diversity of experience in poverty reduction performance over the last few decades and assessing its causes. The other two studies are more methodological in their content, exploring how poverty and related issues can be analysed at the household level using existing survey data that pose difficulties as they do not track the same set of households over time.

Chapter 2 highlights the fact that although growth has improved substantially in most African countries in recent years, poverty across the continent has fallen very little in the aggregate. Poverty reduction performance has varied across countries: there are apparently 'two Africas', one with an ability to reduce poverty and one without. The main argument is that one reason for this difference is rooted in colonial times. Countries with strong smallholder cash crop sectors emerged into independence with broad-based labour-intensive economies supporting a more equitable income distribution conducive to inclusive growth and poverty reduction compared to initially more inequitable mineral resource and large farm based economies. This did not necessarily determine the post-colonial path: many peasant export economies achieved no poverty reduction (often because of little growth), and some mine/plantation economies did achieve poverty reduction. The key reasons for this evolution lie in the motivation and ability of African elites to form pro-poor coalitions, which in some cases were then able to implement policies supporting a pro-poor pattern of growth.

Chapter 3 focuses on pseudo-panel estimation and how it enables the estimation of panel models when only repeated cross-sections rather than panel data are available. It involves the grouping of individual into cohorts and using the cohort means as if they are observations in a genuine panel. The usual assertion is that as long as there are enough individuals within cohorts so that the cohort sample means are a good approximation of the cohort population means then pseudo-panel estimates are consistent, otherwise they may suffer from measurement error bias. We show that

there can be substantial bias arising from the grouping process itself due to the loss of variation and heterogeneity as one moves from the individual to the cohort level. Thus we find many of the common methods used for grouping into cohorts produces inconsistent estimates. We develop some measures for assessing whether the cohort level data contains sufficient variation for pseudo-panel estimates to be consistent, focusing on the variation across cohort means, over time and within cohort groups. We then test the measures empirically and with Monte Carlo simulations, providing useful thresholds that can be used to accept or reject the cohort construction method.

Chapter 4 assesses four different estimation methods of binary response models with individual effects where the data is a time-series of independent cross-sections. We compare Deaton's (1985) linear fixed effects estimator, which is most often used in applied work, to three non-linear estimators. The first is a simplified version of Collado's (1998) Minimum Distance estimator. The other two are based on the fractional response estimators developed by Wooldridge and Papke (2008) which, unlike Collado's estimator, can be used for dynamic models as well. Results from Monte Carlo simulations and from an empirical application indicate the linear estimator is just as good as the other non-linear estimators and is generally more robust to problems arising from the process of creating the pseudo-panel.

Acknowledgements

First and foremost I want to thank my supervisors Oliver Morrissey and Sourafel Girma for their invaluable support, advice, and guidance throughout these last four years. I am especially indebted to Professor Morrissey who encouraged me to pursue this PhD and helped secure the funding for it. I also wish to express my gratitude to my colleagues from B51 with whom I shared an office for the whole duration of my PhD. Their support, friendship, and humour was indispensable in helping to get through the daily grind and occasional stresses that arose during this time.

I gratefully acknowledge the financial support received from the ESRC studentship which made this study possible.

Last but certainly not least, I wish to thank my family for their encouragement, understanding, and patience without which this work could not have been completed.

Table of Contents

Chapter 1: Introduction **1**

Chapter 2: Two Africas? Why Africa's Growth Miracle Barely Reduced Poverty **5**

2.1 The Issues ……6

2.2 A Dichotomy in Colonial Experience ……..……...….8

2.3 The Statistical Evidence ...…….……...…..……..….13

2.4 Empirical Testing and Validation ...……...……..…….....26

2.5 Conclusion ……..……...……...…….37

Appendix to Chapter 2 ...……...…….41

Appendix A ...…….41

Appendix B ...…….48

Chapter 3: Constructing Cohorts to Obtain Consistent Pseudo-Panel Estimates **55**

3.1 Introduction ……..……...…….…….56

3.2 The Importance of Cohort Construction ...……...……...……..58

3.3 Data and Methodology ……..……...…..…….66

3.3.1 Data ...……..……..……..66

3.3.2 Methodology ...……..……..……66

3.3.3 Model and Results from Panel Estimation ...……...72

3.4 Pseudo-Panel Results ...……..……...……76

3.5 Optimal Use of Construction Variables ...……...……...……89

3.5.1 Age and the Width of its Bands .....……...……..…...…….90

3.5.2 Categorical Construction Variables ...……..……...……98

3.6 Assessing Cohort Construction ...……...……..……...….106

3.6.1 Deriving AWAR ...……...……..……...…….106

3.6.2 AWAR and IV-based Alternatives ...……...……...…….115

3.7 Monte Carlo Simulations ...……...……..…..……123

3.7.1 Testing AWAR and its Threshold ...……..……...……..123

3.7.2 Adjusting AWAR for Cell Size ...131
3.8 Alternative Estimation Techniques ...135
3.9 Conclusion ...139
Appendix to Chapter 3 ...142
Appendix A ...142
Appendix B ...144

Chapter 4: Pseudo-Panel Estimation of Binary Response Models 147
4.1 Introduction ...148
4.2 Binary Response Panel Data Models ...151
4.3 Pseudo-Panel Estimators of Binary Response Models ...155
4.4 Monte Carlo Simulations ...158
4.5 Empirical Application ...168
4.6 Allowing for Endogeneity ...175
4.7 Conclusion ...178
Appendix to Chapter 4 ...181

Chapter 5: Conclusion 183

References 190

List of Tables

2.1 Sample Countries and Poverty Performance 1990-2013 ……..……..15

2.2 Population Weighted $1.25-a-day Headcount Poverty …….……19

2.3 $1.25 Poverty in Crop-based vs Resource-based Countries ……...……...23

2.4 Changes in poverty, inequality and output in Crop vs Resource Countries ...24

2.5 Key Factors in Crop vs Resource Countries ..……..……...25

2.6 $1.25 Poverty Performance and PPE ……..……...26

2.7 Determinants of Poverty Performance (full sample, $1.25 weighted) ….....31

2.7b Determinants of Poverty Performance with Additional Time Corrections 32

2.8 Determinants of $1.25 Poverty (additional variables) ……...….....33

2.9 Determinants of £1.25 Poverty Performance (split sample) ……...…….....35

2.10 Determinants of Poverty Performance using Short-term Changes.…..….....36

2.A1 Data Sources and Definitions …….……...41

2.A2 Sample Countries and $1.90 Poverty Performance 1990-2013 ..……...…...43

2.A3 Tests for Difference in Group Means ($1.25) ……....……...44

2.A4 Tests for Difference in Group Means ($1.90) …….....44

2.A5 Determinants of Poverty Performance ($1.90 headcount) ……...45

2.A6 Determinants of $1.90 Poverty (additional variables) ……....…….....46

2.A7 Determinants of $1.90 Poverty Performance (split sample) …….....47

2.B1 Replication of Table 2.7 with Random Effects Estimation ……...….....50

2.B2 Replication of Table 2.8 with Random Effects Estimation …..…….....51

2.B3 Replication of Table 2.9 with Random Effects Estimation …….....52

2.B4 Replication of Table 2.7 with Population Averaged Method …….....52

2.B5 Replication of Table 2.8 with Population Averaged Method …….....53

2.B6 Replication of Table 2.9 with Population Averaged Method …….....54

3.1a Pseudo-panel Estimation in Development Economics …….....70

3.1b Pseudo-panel Estimation of Life Cycles models of Consumption and Labour Supply ….....71

3.1c Pseudo-panel Estimation in General Models of Individual Behaviour ….....71

3.2 Estimation of full model of household consumption …….....74
3.3 Cohort Construction Variables …….....77
3.4a Pseudo-panel estimates based on common cohort specifications …….....84
3.4b Accurate Pseudo-panel estimates …….....85
3.5 Summary statistics of 65 Pseudo-panel estimates …….....85
3.6 Age bands …….....90
3.7 Average Absolute Bias across Age bands …….....90
3.8 Relationship between Estimation Bias and Construction variables …….....93
3.9 Relationship between Estimation Bias and Cohort specifications ….....94
3.10 Re-estimation with 5 Year bands as the baseline comparison ….....96
3.11 Absolute bias across constructional variables …….....99
3.12a Distribution of Household Size across cohort construction variables ….....102
3.12b Distribution of Log Total Assets across cohort construction variables 102
3.12c Distribution of Secondary Occupation across cohort construction variables ……..…… 103
3.12d Distribution of Remittances across cohort construction variables …….....103
3.13 Pseudo-panel estimates from different cohort specifications ….....105
3.14 Correlation between AWAR and weak instrument measures ….....119
3.15 Pseudo-panel estimates passing the threshold tests ….....122
3.16 Simulation Results for Varying the Number of Cohorts …….....127
3.17 Simulation Results for Varying the Degree of Autocorrelation ….....128
3.18 Simulation Results for Varying the Number of Time Periods ….....129
3.19 Simulation Results for Varying the Cell Size …….....130
3.20a Varying Cell Size for CAWAR ….....132
3.20b Varying the Number of Cohorts for CAWAR …….....132
3.21 CAWAR Thresholds for Different Timer Periods …….....134
3.22 Alternatives to Pseudo-Panel estimation ……..….....136
3.23 Alternative techniques for estimating Pseudo-panels ……..……..….....137
3.A1 Summary Statistics of Main Variables …….…….....142
3.A2 Summary Statistics for Age of Household Head ….....143
3.B1 Estimating Table 3.4a using the Simplified Model 145
3.B2 Estimating Table 3.4b using the Simplified Model 145
3.B3 Estimating Table 3.15 using the Simplified Model 146

4.1 Monte Carlo Simulation results for varying σ_x^2 ….....163
4.2 Monte Carlo Simulation results for varying ρ …….…....164
4.3 Monte Carlo Simulation results for varying σ_ς^2 ….....164
4.4 Monte Carlo Simulation results for varying AWAR ….....165
4.5 Monte Carlo simulations comparing Panel and Pseudo-Panel estimates 166
4.6 Panel estimates of Households having above Median Consumption 170
4.7 Pseudo-Panels using 2-year age cohorts, region, location, gender and education ……...…….....174
4.8 Pseudo-Panels using 2-year age cohorts, region, location and education ...174
4.9 Pseudo-Panels using 2-year age cohorts, region, gender and education 175
4.A1 Re-estimation of Table 4.6 with the Simplified Model 181
4.A2 Re-estimation of Table 4.7 with the Simplified Model 181
4.A3 Re-estimation of Table 4.8 with the Simplified Model 181
4.A4 Re-estimation of Table 4.9 with the Simplified Model 182

List of Figures

2.1 $1.25 Poverty Performance over Period 20
2.2 $1.90 Poverty Performance over Period 21
3.1 Distribution of the number of cohorts ……...……....78
3.2a Pseudo-panel estimates from all 65 Cohort Specifications 80
3.2b Pseudo-panel estimates from all 65 Cohort Specifications 81
3.3a Number of Cohorts and Overall Estimation Bias ……...……...88
3.3b Number of Cohorts and Estimation Bias 89
3.4 Distribution of regressors across age of household head 97
3.5a Simulated and Constructed Cohort Estimates 107
3.5b Simulated and Constructed Cohort Estimates 108
3.6 Relative Within Variation and Estimation Bias 110
3.7 Relative Across Variation and Estimation Bias 111
3.8 Across to Within Variation and Estimation Error 112
3.9 AWAR and Estimation Bias ……..…….....114

3.10 F-Statistic and Estimation Bias 116
3.11 First-stage Regression R-squared and Estimation Bias 117
3.12 Shea's Partial R-squared and Estimation Bias 117
3.13 Minimum Eigenvalue Statistic and Estimation Bias ……..…….....118
3.14 Distribution of Autocorrelation Coefficients from Empirical Application ...133
3.A1 Distribution of the Age of the Household Head 143
3.B1 Reproduction of Figure 3.9 using the Simplified Model 146

List of Maps

2.1 Country Classifications 10
2.2 Country $1.25 per day Poverty Performance 17
2.3 Country $1.90 per day Poverty Performance 18

Chapter 1:

Introduction

Analysis of poverty reduction continues to be a vital area of research particularly in Sub-Saharan Africa where poverty levels remain high even after over a decade of impressive growth. Aggregate poverty levels have barely changed since the 1980s in contrast to other developing regions where poverty has fallen by a half or more over the same period. This aggregate picture conceals a wide range of experiences within SSA; some have managed significant reductions in poverty whereas others have seen an increase. This thesis consists of three self-contained studies which connects a diverse set of approaches that can be used to analyse poverty in a more disaggregated and nuanced manner. The first study highlights the cross-country variation in poverty dynamics witnessed in Africa since 1990 and examines the role of colonial history on current poverty performance. The other two studies, rather than conduct further analysis on poverty, details an estimation method outlining how poverty and related issues can be analysed at the household level using existing data. Since the 1980s many household surveys have been conducted in Africa, especially under the World Bank's Living Standards Measurement Survey (LSMS) project, however most surveys are repeated cross-sections (RCS) rather than panel. The problem with RCS is that a different set of households are surveyed in each wave meaning there is no tracking of households across time. This imposes limitations on the type of analysis and modelling that can be conducted. For example one cannot analyse the determinants of households moving out of and falling back into poverty, a key concern for many policymakers.

Chapter 2 explores the nuance required when discussing poverty dynamics across Sub-Saharan Africa by investigating if countries can be grouped into those that reduce poverty and those where poverty rates are persistent. We will test the proposition that a country's colonial heritage influences later performance in reducing poverty by distinguishing crop-based countries, which had an economic structure based around peasant export of cash crops, and resource-based countries with a legacy of settler capitalism where land was appropriated by European settlers for production of exports

(often via plantations) or the mining of mineral resources. We test a number of other factors that may influence the success of a country in reducing poverty, namely: growth, the level of pro-poor expenditure (a proxy for redistributive policies), inequality and the level of democratisation.

Chapter 2 gives a broad overview of the importance of historical and political factors on current poverty dynamics in Africa, but does not identify factors that made growth pro-poor or specific policies that can contribute to reducing poverty at the household level. In order to answer this question one needs to look more at the household and individual level to identify the causes of and constraints to successful poverty reduction. One way would be to look at a country which has successfully reduced poverty over this period, such as Uganda, to see which households escaped poverty and remained out of it, which households escaped and fell back in, and which households never managed to escape. In this way one could produce policy proposals that target the specific conditions that aid the movement out of poverty as well as reduces the risk of falling back into it. Examples of this would be to look at the role of agricultural production, urban migration, and health shocks. Continuing with the Ugandan example, there are detailed household surveys covering the whole period where poverty has been falling (from the early 1990s to the present) which could potentially be used for this type of analysis. The problem is the data are RCS and not panel, thus the analysis and modelling that can be conducted is limited to cross-sectional estimation. This means tracing movements in and out of poverty is not possible, one cannot account for fixed effects which may be necessary to control for unobserved factors that are correlated with explanatory variables (such as innate ability of household members or their social networks), and using dynamic specifications such as lagged variables. While all this can be done with panel data Uganda only has it for a relatively short period of time with the first wave beginning in 2005/06, meaning such analysis would miss the major changes that occurred in the first decade of their successful poverty reduction program. Data limitations of this type are typical for most African countries and one would face similar problems whether we wanted to analyse the successful cases or the unsuccessful ones.

One could address this issue by producing the requisite panel datasets. Not only are there great costs and difficulties associated with this, such as the problem of attrition which only worsens the longer the time period one requires, but it would still not help draw lessons from the changes that occurred over the 1990s and 2000s. Chapters 3 and 4 extend methods to utilise the existing RCS data, which has been collected from the early 1990s onwards, and still allows for panel-type analysis such as tracing movements in and out of poverty. The method we use is pseudo-panel estimation first proposed by Deaton (1985) which, due to the problems we have just highlighted, has increasing been used in applied studies focusing on developing countries in general (not just in Africa) and particularly for analysing household poverty dynamics. Pseudo-panel estimation involves the grouping of households (or individuals) into cohorts based on time invariant characteristics such as gender or year of birth. Then the cohort means are used as if they are observations in a genuine panel. In this way it does not matter if different households are surveyed in each wave as long as the cohorts that households belong to can be traced over time. In effect, one can think of cohorts as being representative households with which estimation is conducted.

Chapter 3 explores exactly how representative the cohorts are of the underlying households and how this affects the regression estimates. This question has not been sufficiently addressed in the literature and many of the common methods of grouping into cohorts lead to heavily biased estimates which are often worse than using simple OLS. We will explore alternative methods for cohort grouping and develop tests that can identify whether a grouping method is likely to produce unbiased estimates. The test will combine and express as a single value three crucial types of variation present in the cohort level observations; the level of inter-cohort heterogeneity (how much genuine variation there is across the cohort means), the level of intra-cohort homogeneity (how similar households within a cohort are), and the level of time variation in the cohort means (the degree of serial correlation at the cohort level). We also discuss some of the other practicalities involved when using grouping variables to create cohorts.

The analysis conducted in Chapter 3 will be based on a linear fixed effects model which is relevant to many empirical applications that use household surveys and other

data found in RCS form. However if one is interested in analysing poverty or movements in and out of poverty (as well as a host of other important models looking into issues such as unemployment, health outcomes, and educational attainment) a linear model may not be suitable as it does not account for nonlinearities caused by the boundedness of the binary dependent variable. Chapter 4 will explore the estimation of binary response models under a pseudo-panel setting. The performance of four possible estimators will be assessed: the linear estimator from the previous chapter, and three nonlinear estimators adapted from the literature on nonlinear panel estimation for a pseudo-panel setting. The four estimators will be tested using Monte Carlo simulations and an empirical application again using Ugandan data.

The aim of the thesis is to develop tests and explore methods for greater use of pseudo-panel techniques in SSA countries where available surveys are RCS rather than panels. We believe there is great potential for using pseudo-panel estimation given the data limitations for analysing key issues such as poverty not just for African countries but in most of the developing world. However too little attention at present is given to the process of cohort construction in existing studies making it difficult to assess their validity. There is a wealth of RCS data that has been collected over the last three decades, particularly in Africa, but their use is often constrained by their inability to trace households/individual over time. Pseudo-panel estimation has the potential to use such data to address crucial policy matters that would otherwise not be possible.

Chapter 2:

Two Africas? Why Africa's Growth Miracle Barely Reduced Poverty[1]

Abstract

Although growth has improved substantially in most African countries in recent years, poverty across the continent has fallen very little in the aggregate. Poverty reduction performance has varied across countries: there are apparently 'two Africas', one with an ability to reduce poverty and one without. The main argument of this paper is that one reason for this difference is rooted in colonial times. Countries with strong smallholder cash crop sectors emerged into independence with broad-based labour-intensive economies supporting a more equitable income distribution conducive to inclusive growth and poverty reduction compared to initially more inequitable mineral resource and large farm based economies. This did not necessarily determine the post-colonial path: many peasant export economies achieved no poverty reduction (often because of little growth), and some mine/plantation economies did achieve poverty reduction. The key reasons for this evolution lie in the motivation and ability of African elites to form pro-poor coalitions, which in some cases were then able to implement policies supporting a pro-poor pattern of growth.

[1] This chapter was written in combination with Paul Mosley (University of Sheffield) and Oliver Morrissey. A CREDIT Working Paper version can be found here: http://www.nottingham.ac.uk/credit/news/papers/1601.aspx

1 The Issue

After decades of economic stagnation or even decline, since the late 1990s sub-Saharan Africa (SSA, or simply Africa hereafter) has recorded more than a decade of impressive growth, averaging 5-6 per cent per year (Devarajan, 2013: S9) and generating talk of 'African cheetahs' to rival the Asian tigers (Martins 2013). In terms of reducing poverty, however, achievements have been far less impressive. Poverty in SSA (at $1.25 a day purchasing power parity) is certainly moving in the right direction, but has declined by only three percentage points, from 51% to 48%, since 1981, during a period when the developing world as a whole has more than halved poverty, from 52% to 21%.[2] For this reason, poverty in Africa will not have fallen sufficiently to meet the Millennium Development Goals, and in particular the primary target of halving of dollar-a-day poverty by 2015 (UNDP, 2014: xiii).

Country examples of successful and sustained poverty reduction in Africa are not hard to find. Ghana and Uganda have managed to halve poverty over the last twenty years (Aryeetey and McKay, 2007; Okidi *et al*, 2007) and there is evidence of long-term declines in poverty for Botswana, Burkina Faso, Cameroon, Ethiopia, Guinea, Senegal and Tanzania amongst others. With the exceptions of Ethiopia and Tanzania, however, none of these relative success cases is a large and populous country, and this is one factor reducing SSA aggregate performance; large countries such as Kenya, the Democratic Republic of Congo (DRC) and Nigeria have not significantly reduced poverty. Could it be that in terms of poverty dynamics there are two separate Africas, in one of which growth is inclusive and broad-based, hence poverty-reducing, whereas in the other it is not? This is the proposition demonstrated and explored in this paper.

There is considerable disagreement on the actual trends in poverty reduction in Africa. The most optimistic view alleges that aggregate poverty in Africa is falling faster than the official statistics indicate. Inferring poverty trends from the consumption component of GDP, as done by Pinkovskiy and Sala-i-Martin (2010), is however

[2] During this period, poverty in East Asia fell from 61% to 13%, in South Asia from 66% to 26%, and in Latin America and the Caribbean from 12% to 6% (Olinto and Uematsu, 2013); see also Chen and Ravallion (2010).

likely to overstate poverty reduction (McKay, 2013; Ravallion, 2000). Nevertheless, carefully analysed household surveys that are more informative on poverty trends (Deaton 2005) are also generally optimistic. McKay (2013) reviews evidence from household survey data for 25 SSA countries to demonstrate significant achievements in poverty reduction, albeit with variation across countries and at a lower rate than claimed by Pinkovskiy and Sala-i-Martin (2010).

The pessimistic view argues that because national accounts data in SSA are of such low quality GDP is measured very inaccurately and poverty is probably higher than estimated (Jerven, 2013). Although the statistical difficulties are valid the inferences drawn are overstated: there is little doubt that the majority of SSA countries have experienced more than fifteen years of significant growth; there may be inaccuracies in GDP but relative levels and growth rates within SSA are likely to represent meaningful experience. The pessimistic view is also sceptical of household survey data given practical problems in, for example, defining and measuring 'the household' and informal employment (Jerven and Johnston, 2015). Such limitations, however, do not render the data uninformative and there is robust evidence of long-term poverty reduction in several countries.

Section 2 presents the argument that variations in colonial experience created some countries where growth had the potential to be inclusive, given the better opportunities for the poor because production was based on the development of labour-intensive smallholder agriculture. The prospects for inclusive growth were less in other countries whose production was based on more capital-intensive mines and plantations giving rise to narrow-based growth that excluded most of the population. Section 3 presents evidence for a dichotomy in poverty reduction performance that reflects the varying success of efforts to build on, or where necessary counteract, those initial conditions. Our notion of 'two Africas' is based on a distinction between a group of SSA countries that have achieved poverty reduction through inclusive growth and another group with little or no poverty reduction despite growth. The hypothesis that the colonial experience helps to explain differences in poverty reduction is tested with a parsimonious model in Section 4 (with further analysis in Appendix B). Section 5 discusses the findings and concludes.

2 A Dichotomy in Colonial Experience

The notion that there may be distinct types of African country in terms of poverty reduction performance is not new. Bowden, Chiripanhura and Mosley (2008) argued, in relation to a small sample of African countries, that the division between 'two Africas', one with a potential for inclusive development and one much less so, has its roots in colonial times, and that this has relevance to the poverty reduction strategies which have a chance of being effective today (see also Bowden and Mosley, 2012). During the late nineteenth century, two alternative strategies were used to try and make colonies economically viable (Hancock, 1943; Myint, 1976). The strategy of *settler capitalism*, allowing European settlers to appropriate agricultural land for the production of export crops (and often also mining), applied to South Africa, Zimbabwe and parts of Angola, Mozambique, DRC, Kenya and Zambia, tended to be associated with high levels of initial inequality and capital-intensity. This category is similar to what Mkandawire (2010) characterises as either labour surplus or concession company colonialism, in which the colonial power exercised considerable economic control and inequality was relatively high. The alternative, *peasant export development,* found for example in Senegal, Ghana, Sierra Leone, Nigeria, Uganda and Tanzania, provided more opportunities for Africans to develop skills and even to achieve promotion into administration. In this strategy, agricultural production was largely in the hands of indigenous African smallholders (in some cases settlers were prohibited from owning land), although marketing was controlled by the colonial government, and an infrastructure developed that was adapted to exports of smallholder crops. This is similar to what Mkandawire (2010) characterises as cash crop economies, where a relatively large informal sector developed and inequality was somewhat lower.[3]

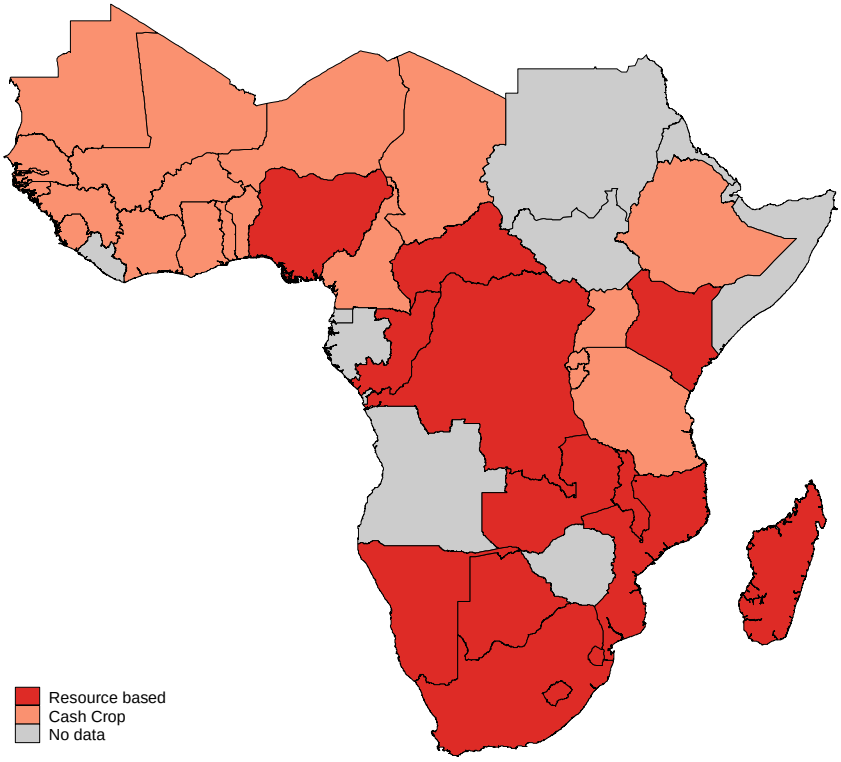
The fundamental point is that smallholder agriculture was important in its own right as a process for creating broad-based development, given relatively strong linkages with the rest of the economy, and also generated a politics in which advancement of low-income Africans was easier to achieve. Even by the late 1920s, substantial

[3] Note that whereas Mkandawire (2010) was concerned with the effect of this colonial heritage on the evolution of institutions that influenced the tax capacity of the state our concern is with the influence on views towards distribution.

differences in infant mortality, real wages and other measures of well-being had emerged between settler economies and peasant economies (Bowden and Mosley, 2013: figure 13.2). Peasant economy status provided a legacy in the shape of a relatively equitable initial distribution of income and assets which eventually provided several African countries with a valuable platform for an inclusive development strategy. In contrast, mining and plantation-based economies were based on sectors with relatively weak linkages to the rest of the economy. This fostered a more exclusive pattern of growth and created incentives for rent-seeking and a politics that was not pro-poor.

Colonial heritage is only a broad categorization and in many countries there was no clear dichotomy. For example, when Tanganyika (now Tanzania) was in German hands from the late 19th century until World War 1 there was an influential settler community in the North keen to promote a plantation economy in conflict with a Governor who wanted to promote African smallholder production in the centre. The result was a compromise in effect if not intent (Iliffe, 1969), a hybrid economy such that Tanzania eventually emerged as more of a peasant than a settler economy, and this was reinforced by the post-independence policies of Julius Nyerere. The classification of other countries changed over time; Ruanda-Urundi is classified by Mkandawire (2010) as among the concession company colonies but post-independence Rwanda and Burundi developed as crop-based economies. More generally, post-independence development was affected by the timing of new discoveries of mineral resources in many countries, such as oil in Nigeria. The proposed dichotomy should be viewed as a distinction between production structures and policies likely to yield inclusive or exclusive development. In some SSA countries, which will be termed *crop-based*, smallholder agriculture production of food and cash (export) crops was important and this provided a foundation for political institutions that favoured broad-based development. The other countries will be termed *resource-based*, where mining and plantation or large scale export crop production was far more important and institutions emerged to reflect this enclave style of development. Map 2.1 illustrates the distribution of the two types of economy.

Map 2.1: Country Classification



Note: Classification adapted from Mkandawire (2010) by combining labour reserve and concession companies into Resource-based and reclassifying some countries that have changed significantly since independence (Nigeria classified as Resource-based; Burundi and Rwanda as Cash-crop). 'No data' designates countries omitted from the analysis due to data limitations.

The colonial inheritance of a crop-based economy is neither a necessary nor a sufficient condition for a successful poverty-reduction strategy. Several crop-based economies, under the stress of subsequent resource discoveries, civil conflict and/or governance deficiencies, experienced failed development (especially growth) strategies after independence and had dissipated the initial potential by the 1980s, Nigeria and Côte d'Ivoire being prominent illustrations. In contrast, some resource-based economies were able to avoid the natural resource, capital-intensive trap and develop effective inclusive poverty reduction strategies; Botswana may be the best example (Poteete, 2009) but Namibia is another. We are not claiming that the nature of the colonial experience as described here is deterministic but rather that it created initial conditions that either increased the likelihood of inclusive policies emerging that were favourable to poverty reduction (the relative political and economic importance of rural, smallholder producers in crop-based economies supported broad-based politics) or made exclusive rent-seeking politics more likely (in resource-based

economies where particular groups had preferential access to economic and political influence). Two factors were particularly important in determining how these initial conditions evolved: the nature of post-independence leadership and the timing of significant new resource discoveries.

The contrast between Ghana and Nigeria illustrates the impact of resource discoveries. Both originated as crop-based economies in which smallholder exports (of cocoa and palm oil) served as the mainspring of economic growth and provided the buffer of low initial inequality deriving from the wide diffusion of gains from those exports. However, from the 1950s onward after the discovery of oil, Nigeria shifted to a resource-based economy following long periods of military government and political turbulence, especially the civil war in the mid-1960s. The conventional wisdom is that in Nigeria the political system was not able to resist, and indeed amplified, the effects of the 'natural resource curse'. Nigeria's oil resources generated Dutch disease, promoted a capital-intensive pattern of production which was bad at generating new jobs, and stifled rather than encouraged smallholder agriculture where most of the poor were based (Bach, 2004; World Bank, 2013). Although oil provided the government with significant revenues for investment and spending these were not used to finance public spending on smallholder agriculture, the mainstay of most poor people's subsistence (Mogues *et al*, 2008), or investment in education. The rent-seeking induced by resource wealth ensured that growth in Nigeria has not been inclusive.[4]

Post-independence Ghana was politically a relatively stable country with a strong smallholder cocoa sector, even under the regimes of President Flight-Lieutenant Jerry Rawlings. It was one of the first African countries to abandon single-party, quasi-authoritarian politics in favour of genuine multi-party democracy in the 1992 elections won by the National Defence Council (NDC), the more rural-oriented and less business-oriented party, which initiated political organization of the poor with Rawlings elected as President. The focus on educational policies was a key instrument to capture the loyalty of poor voters; 'after taking power Rawlings and his …

[4] Ghana, since 2009, has of course become an oil exporter and there has been much anxiety about the potential distributional and political consequences of this, some of it explicitly focussed around the question of how to avoid the consequences of the Nigerian precedent. For an early discussion of the fiscal management of offshore oil in Ghana see Kopinski *et al* (2013).

[government] organized, and served, the poor' (Kosack, 2012: 200). Despite this political orientation towards rural areas, the decline in headcount poverty in Ghana over the course of the 1980s and 1990s largely benefited workers (including farm workers) and the self-employed and was more pronounced in urban than rural areas (Nsowah-Nuamah *et al*, 2010); the benefits were least in subsistence farming areas of the north (Porter, 2003; World Bank, 2011; Olinto and Uematsu, 2013). This may disguise the benefits for commercial farmers and cocoa smallholders who were above the poverty line by the mid-1980s. The key elements in the reform process from the 1980s did benefit agriculture: tax reform and tariff reduction; liberalisation of the exchange rate regime; and privatisation of statutory marketing boards resulting in large increases in the on-farm price received by producers.

Post-independence politics were instrumental in the way in which distributional politics evolved in east Africa. Conditions were most favourable in Tanzania: ethnic divisions were not as pronounced as in other countries, in part because the widespread adoption of Swahili supported unity in nationalist and post-colonialist eras (Lonsdale, 2012: 40), and Nyerere successfully promoted an inclusive politics. Kenya, in contrast, was more ethnically divided and this was exploited by political leaders to create a divisive form of ethnic patriotism (Lonsdale, 2012). White settler interests were concentrated in the Rift Valley and in the post-independence period Kenyatta supported resettlement of Kikuyu there, forging an (ultimately fragile) alliance with Kalenjin so the ethnic groups in the leadership captured control of resources. 'The struggle to control state power has led to state-sponsored ethnicity [using] state institutions to create both class interests and ethnic domination by a system of reward and exclusion' (Ogut, 2012: 122). Political stability in Uganda only lasted until the early 1970s and it was not until 1986 that fourteen years of political turbulence and five years of guerrilla war came to an end. The victory of Yoweri Museveni and the National Resistance Movement (NRM), Uganda's only legitimate political party, permitted a 'new beginning' and initially had priorities for restoration of the rule of law, the economy and infrastructure; prior to the early 1990s poverty and poverty reduction were scarcely considered. The three key policy measures which caused poverty to almost halve between 1992 and 2000 were, as in Ghana, agricultural reform including the removal of taxes on exports (coffee and cotton – the bulk of which were

produced by low-income smallholders); the prioritising of pro-poor expenditures including primary health and education, rural infrastructure and smallholder agriculture; and the pursuit of universal primary education (Mosley, 2012). An important element in the allocation of resources in Uganda has been a process of decentralization to local councils and allowing an increasing role for non-governmental organizations (Barr and Fafchamps, 2006). Advocates for the poor were incorporated into the machinery of government (through the dominant NRM party) to serve as a buttress against protest, and the whole process was reinforced by the boost to rural smallholder income provided by agricultural liberalization.

The colonial experience was rarely a determinant of distributional policy in the 1960s, although it was a factor in how post-independence politics evolved (inclusive in countries like Tanzania but divisive in countries like Kenya). In most African countries the 1970s and much of the 1980s was a period of political and economic instability and at best low growth (and high poverty). Following adjustment in the 1980s (at the instigation of donors) more stable policy regimes began to emerge in the 1990s. This is where we take up the story in the next two sections; whether the colonial legacy was converted into a durable pro-poor development strategy depended on whether a pro-poor coalition could be assembled which had the vision, the technical capacity and the motivation not only to break out of the low-income trap – which most of Africa appeared to be in by the 1970s – but also to make a dent in poverty levels through growth supported by an effective pro-poor policy framework.

3 The Statistical Evidence

Alternative ways of measuring poverty give rise to alternative estimates. South Africa is perhaps an extreme example where apparently the same source, national surveys, can give very different estimates of levels and trends in poverty.[5] Our main analysis

[5] South Africa has multiple and often conflicting sources of data on poverty (Mosley, 2013, Appendix part (iii)). Leibbrandt et al (2010: tables 2.10 and 2.11) suggest that between 1993 and 2008 poverty declined very slightly from 56% to 54%, although it increased within the Coloured

uses the World Bank Povcal $1.25 a day measure as that covers the most countries for the longest period using standardized poverty lines (data sources and definitions are in Appendix Table 2.A1). Poverty estimates and trends using the more recent $1.90 poverty lines are presented in Appendix Table 2.A2. In countries which have experienced periods of conflict data are often missing and the period of measurement is crucial; Rwanda has reduced poverty dramatically since 1996 but the devastating effect of the 1994 genocide means that on the World Bank's Povcal measure poverty was at the same level in 2011 as it had been in 1986 (national poverty line poverty fell much more significantly in the 2000s and suggests a better performance).

Tanzania is a good example of a country where alternative data sources provide differing evidence. The $1.25 a day POVCAL data suggests that poverty fell from a headcount of 72% in 1992 to about 44% in 2012, implying that poverty fell by almost 40 per cent over some 20 years. The national poverty line derived from the Tanzania Household Budget Surveys suggests much lower levels and declines in consumption poverty: over the same 20 years poverty fell by only 27 per cent, from 38.6% in 1991/92 to 28.2% in 2011/12. Arndt *et al* (2015) show that differences in the price deflator used (and of course the purchasing power parity conversion) explain much of the difference between consumption poverty rates derived from national survey data compared to POVCAL estimates. The major reason why growth since 2000 did not translate into commensurate poverty reduction was the lack of growth in real household consumption; the relatively high GDP growth since 2000 was in components other than private consumption, in particular the mining sector (Arndt *et al*, 2015: 4). The main results are based on the internationally comparable POVCAL data and we do not include estimates using national poverty line due to insufficient data, with only 12 countries having poverty observations over a period of 10 years or more since the 1980s.

population. World Bank PovCal data, using a $1.25 per day poverty line, suggests a much greater reduction and much lower level (Table 2.A2).

Table 2.1: Sample Countries and Poverty Performance 1990-2013

Country	Period	Sur-veys	% Δ GDP pa	$1.25 Poverty Headcount			Elast-icity
				Initial	Current	% Δ pa	
Gambia	1998-2003	2	4.1	65.6	33.6	-9.75	-2.4
Congo, Rep.	2005-2011	2	5.4	54.1	32.8	-6.56	-1.2
Chad	2003-2011	2	10.1	62.0	36.5	-5.13	-0.5
Cameroon	1996-2007	3	4.0	47.4	27.6	-3.80	-0.9
South Africa	1993-2011	6	3.2	24.3	9.4	-3.40	-1.1
Swaziland	1995-2010	3	2.6	78.4	39.3	-3.33	-1.3
Namibia	1993-2010	3	3.8	49.1	23.5	-3.06	-0.8
Botswana	1986-2009	4	5.8	35.6	13.4	-2.71	-0.5
Guinea	1991-2012	5	3.3	93.7	40.9	-2.69	-0.8
Ethiopia	1995-2011	4	7.5	63.2	36.8	-2.61	-0.3
Mali	1994-2010	4	5.5	85.9	50.6	-2.57	-0.5
Burkina Faso	1994-2009	4	5.8	71.2	44.5	-2.50	-0.4
Ghana	1988-2006	5	4.7	50.6	28.6	-2.42	-0.5
Senegal	1991-2011	5	3.5	65.7	34.1	-2.41	-0.7
Niger	1992-2011	5	3.2	72.8	40.8	-2.31	-0.7
Mauritania	1987-2008	6	1.4	41.3	23.4	-2.06	-1.5
Tanzania	1992-2012	4	5.3	72.0	43.5	-1.98	-0.4
Uganda	1989-2013	8	6.8	70.8	37.8	-1.94	-0.3
Mozambique	1996-2009	3	7.7	80.6	60.7	-1.90	-0.2
C. African Rep.	1992-2008	3	2.5	83.1	62.8	-1.53	-0.6
Malawi	1998-2010	3	4.1	83.2	72.2	-1.11	-0.3
Weighted Average 'High Performers'			**5.5**	**62.2**	**36.8**	**-2.64**	**-0.6**
Sierra Leone	1990-2011	3	2.4	62.9	56.6	-0.47	-0.2
Togo	2006-2011	2	3.5	53.2	52.5	-0.26	-0.1
Burundi	1992-2006	3	-0.5	84.3	81.3	-0.25	0.5
Lesotho	1987-2010	5	4.1	57.2	56.2	-0.08	0.0
Rwanda	1985-2011	4	4.8	63.3	63.0	-0.01	0.0
Benin	2003-2012	2	3.7	47.3	51.6	1.00	0.3
Weighted Average 'Uncertain'			**2.9**	**62.4**	**61.7**	**0.06**	**+0.1**
Nigeria	1986-2010	5	4.7	54.1	62.0	0.62	0.1
Kenya	1992-2005	4	2.5	38.4	43.4	0.99	0.4
Zambia	1991-2010	8	4.6	61.1	74.3	1.14	0.2
Guinea-Bissau	1991-2002	3	1.1	41.3	48.9	1.67	1.5
Madagascar	1993-2010	6	2.8	67.1	87.7	1.81	0.6
Côte d'Ivoire	1985-2008	9	1.7	8.1	35.0	14.51	8.7
Weighted Average 'Deteriorating'			**4.0**	**49.5**	**59.7**	**1.84**	**+0.9**

Note: Countries classified as crop-based are shaded in grey; others are resource-based. Period refers to the years between the first (initial) and final (current) poverty observation; Surveys is number during the period; '% **Δ** GDP pa' (GDP growth in % annual average) and poverty '% **Δ** pa' (change poverty % annual average) are calculated over this period; poverty elasticity is % **Δ** poverty/% **Δ** GDP (Burundi

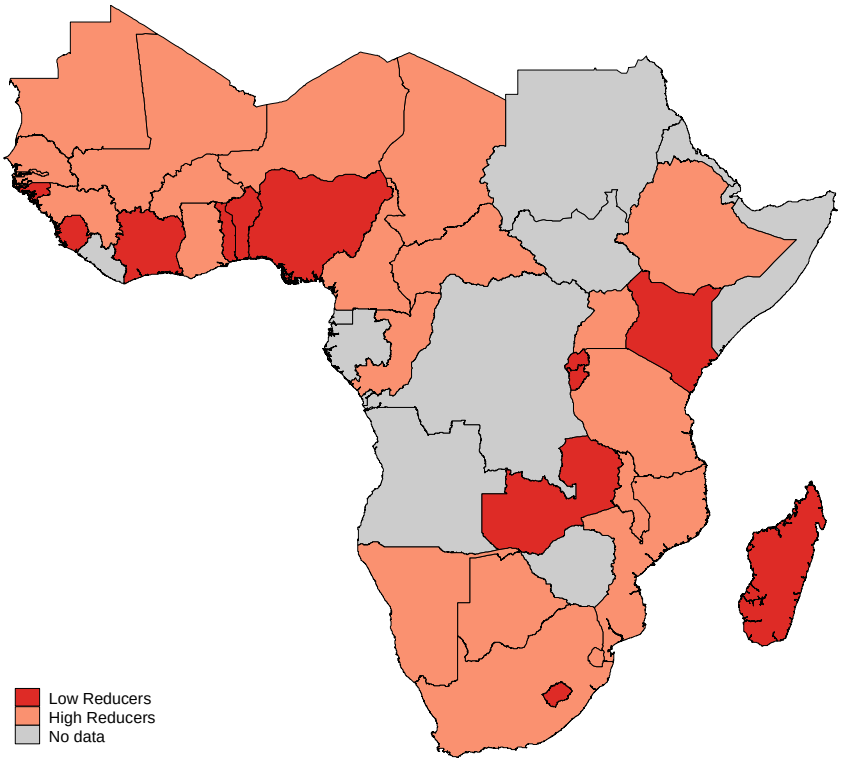
not signed because growth and poverty declined). The 'High Performers' reduced poverty; poverty increased in the 'deteriorating' countries; others are classified as 'uncertain' because there are only a couple of recent observations or inference depends on the period or poverty measure chosen (see Appendix Table 2.A2); means are weighted by population.
Source: World Bank, POVCAL database; see Appendix Table 2.A1.

Table 2.1 reports poverty trends at the $1.25 poverty line between the first survey closest to 1990 and the most recent survey for 33 African countries; 21 appear to have sustained reductions in poverty of at least 1% per annum on average (although in some cases over only a short recent period); six experienced increasing poverty and for another six the change was negligible or data are ambiguous (e.g. trend varies by poverty line used). In general, countries with higher growth reduced poverty by more (Dollar and Kraay, 2002) but this was not consistently the case. The poverty elasticity of growth is variable: for the high performers (who reduced poverty) it is generally between -0.2 and -0.8, but -1.0 or better for six countries; in six countries (deteriorating) poverty increased - Côte d'Ivoire (3.2) and Guinea-Bissau (1.5) have the highest (adverse) values. The majority of high performers (13/21) are classed as crop-based (as are most uncertain countries) and the majority deteriorating are classed as resource-based. Map 2.2 distinguishes those countries that have sustained reductions in poverty since 1990 (high reducers) and those that have not (low reducers, uncertain or deteriorating) for the $1.25 measure and illustrates the tendency for the crop-based countries to perform better than resource-based economies. Map 2.3 presents similar results for $1.90 poverty (Appendix Table 2.A2); the only changes are that Sierra Leone and Lesotho become high reducers whereas Malawi becomes a low reducer (uncertain).

The countries that reduced poverty (high performers) on average had higher initial poverty and higher annual average growth than the countries in which poverty increased (deteriorating) and achieved more significant reductions with a better elasticity (Table 2.A3). The countries classed as uncertain were not statistically different from improvers in terms of initial $1.25 poverty but experienced significantly lower growth and poverty reduction; whereas they initially had higher poverty than deteriorating countries there was no statistical difference in growth and they outperformed in terms of reducing poverty (only a weakly significant difference).

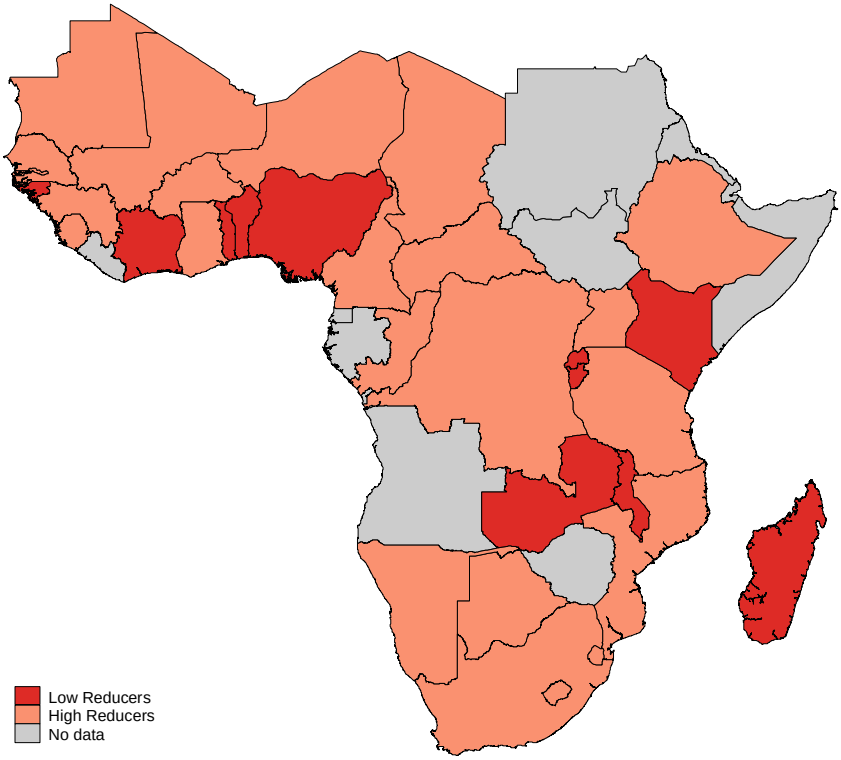
High reducers statistically outperformed low reducers (uncertain and deteriorating) in terms of growth, poverty reduction and elasticity whether using $1.25 (Table 2.A3) or $1.90 (Table 2.A4) poverty.

Map 2.2: Country $1.25 per day Poverty Performance



Note: 'High Reducers' are countries in which poverty clearly decreased (High Performers in Table 2.1); 'Low Reducers' either had small reductions but conflicting data (Uncertain in Table 2.1) or poverty increased (Deteriorating in Table 2.1). 'No data' designates countries omitted from the analysis due to data limitations.

Map 2.3: Country $1.90 per day Poverty Performance



Note: As for Map 2.2.

Table 2.2 provides summary data on poverty reduction in SSA since 1990 (mostly in the 2000s). Panel A covers all countries with data in any period (full sample) and the set of 26 countries for which we have observations close to the years 1990, 2000 and 2010. Panel B splits the set of 26 countries into countries who were able to significantly reduce poverty to almost half the 1990 level by 2010 (high performers) and those that were not (low performers). This gives a more nuanced picture of poverty reduction performance in SSA than the aggregate statistics in Panel A. One can now explicitly see the 'two Africas' phenomenon that this paper is highlighting: two groups of countries start off with similar levels of poverty (albeit higher in the high performers) but then fall into two divergent paths in reducing poverty over the next two decades. The divergence starts to take shape in the 1990s with high performers reducing poverty by almost eight percentage points whereas the low performers saw an average increase in poverty of four percentage points. Over the 2000s, the high performers were able to accelerate the pace of poverty reduction and reduce average poverty levels by a further 17 percentage points whereas the low performers had a further slight increase in poverty. Note also that for both groups the standard deviation

declined over time. Thus by 2010 high performers not only managed to almost halve average poverty from the levels seen in 1990 but also had almost half the level of poverty of the low performers. This divergence is further highlighted by the fact that the high performers initially contained the country with the highest level of poverty in 1990 but by 2010 the highest rate of poverty amongst the high performers was only just above the average poverty rate among low performers. In general the low performers are more populous as the nine countries accounted for about one third of total SSA population, whereas the 17 high performers accounted for just below 43% of SSA population.

Table 2.2: Population Weighted $1.25-a-day Headcount Poverty

Panel A	**Full Sample**			**Consistent set of 26 Countries**		
	1990	**2000**	**2010**	**1990**	**2000**	**2010**
Mean	58.0	56.1	48.9	58.0	55.6	46.6
Std. Dev.	17.5	18.1	21	17.6	18.3	18.1
Min	18.8	21.2	6.1	18.8	21.2	11.5
Max	93.7	86.4	87.7	93.7	86.4	87.7
N	27	32	36	26	26	26
Population	383.2	533.3	819.6	382.2	501.1	650.1
% SSA	75.5	80.3	94.8	75.3	75.4	75.2

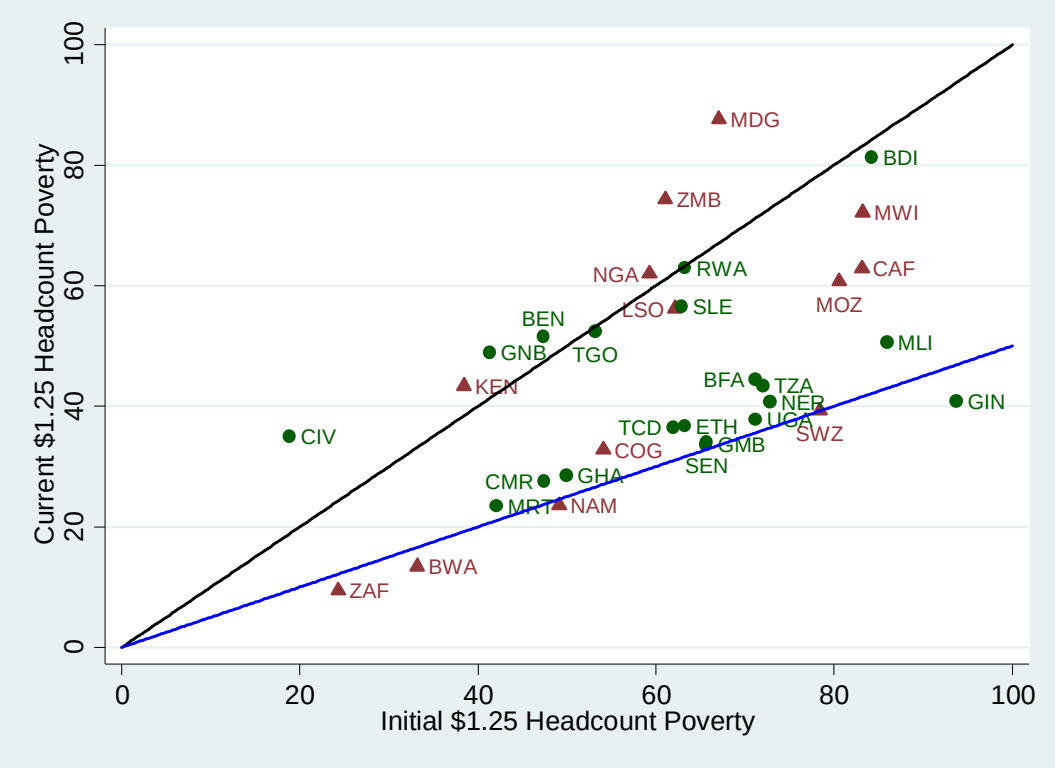
Panel B	**Low Performers**			**High Performers**		
	1990	**2000**	**2010**	**1990**	**2000**	**2010**
Mean	55.4	59.5	60.6	60.1	52.6	35.8
Std. Dev.	14.7	17.1	13.7	19.8	19.3	13.2
Min	18.8	29.9	35.0	24.3	21.2	11.5
Max	84.3	86.4	87. 7	93.7	84.2	62.8
N	9	9	9	17	17	17
Population	169.0	217.2	281.7	213.2	283.9	368.4
% SSA	33.3	32.7	32.6	42.0	42.7	42.6

Notes: High Performers as for Table 2.1, Low Performers combines Uncertain and Deteriorating from Table 2.1; poverty averages based on surveys closest to the indicated year; *N* is number of countries in the particular sample; *Population* is sum of the *N* countries; *% SSA* is sample as a percentage of total SSA population.

This distinction between the 'two Africas' is illustrated in Figure 2.1. The 45^{o} line indicates countries that have not reduced poverty (in any countries above the 45^{o} line, poverty increased) while the lower line represents countries that have reduced poverty by about half since the earliest and most recent survey observation. Crop-based countries are more likely to be clustered around the lower line; the main exceptions are crop-based countries that were initially small and poor (such as Benin and Togo) and/or experienced significant instability (Burundi, Rwanda and Sierra Leone). In

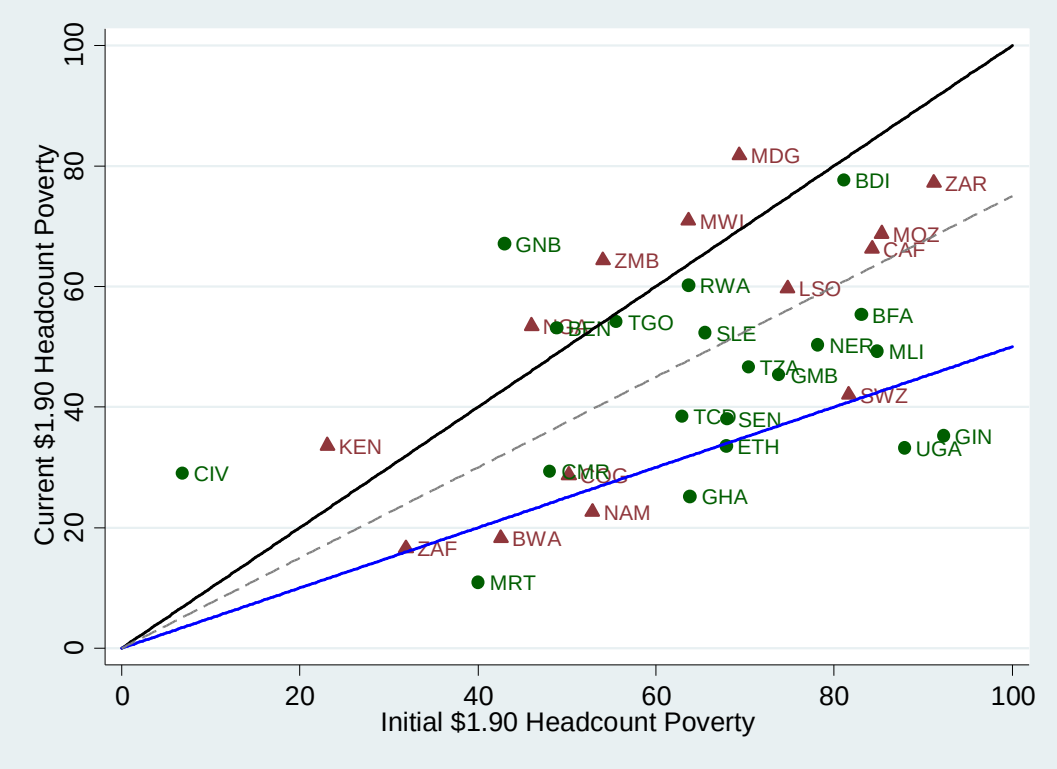
contrast, resource-based economies tend to be around the upper line; the exceptions are in Southern Africa (Botswana, Namibia, South Africa and Swaziland) and those that had relatively good institutions. Obviously, there is a wide dispersion around both lines and it would be misleading to claim a high correlation (the analysis in the next section shows that the differences are significant). The figure also highlights another interesting difference between crop- and resource-based economies in terms of initial poverty. The poverty reduction of resource-based economies appears to be dependent on initial poverty levels: countries with initially low poverty levels were able to further reduce poverty by about a half, whereas those with high initial poverty (above a 40% headcount) struggled to reduce poverty (with the exception of Swaziland) or even saw an increase. For crop-based economies initial poverty levels do not appear to affect subsequent poverty reduction performance. Figure 2.2 presents comparable information for the $1.90 poverty line (the dotted line represents a 25% reduction in poverty): although the countries are more spread out the broad pattern is the same.

Figure 2.1: $1.25 Poverty Performance over Period
(1990-2013 or most recent observation)



Notes: Countries designated ● are classified as crop-based whereas those designated Δ are resource-based. The upper (45°) line is no change in poverty, the lower line indicates a 50% reduction between the first and last survey measure. The basic pattern is unaltered if countries with very low data coverage are excluded. A comparable pattern is found using $2 a day poverty and reduction by 25%.

Figure 2.2: $1.90 Poverty Performance over Period
(1990-2013 or most recent observation)



Notes: Countries designated ● are classified as crop-based whereas those designated Δ are resource-based (raw data in Table 2.A2). The upper (45º) line is no change in poverty, the lower line indicates a 50% reduction between the first and last survey measure. The dotted line is poverty reduction by 25%.

Table 2.3 gives further evidence on the relation between colonial heritage and poverty reduction. Most of the poverty reduction since 1990 has come from crop-based countries and this is more pronounced for the full sample (Panel B) compared to the 26 country sample (Panel A). The table also highlights differing trends within crop-based and resourced-based economies. Not only is poverty falling on average in crop-based economies but poverty levels are also converging over time as shown by the declining range and standard deviation, particularly the unweighted standard deviation which better captures variation between countries. On the other hand, resource-based economies are characterised by average poverty falling little and poverty levels across countries diverging over time, a trend noticeable in Figure 2.1. Although crop-based economies experienced much larger poverty reduction over the period they also started off with higher initial poverty and by 2010 had obtained levels below resource-based economies. Panel B, which contains the full sample of countries, suggests poverty

levels for resource-based economies are under-reported in the restricted sample as some of the poorest and most populous countries are excluded. Using the full sample for 2010 one can see that poverty levels are significantly lower in crop-based countries compared to resource-based. This is largely a consequence of what transpired over the 2000's, where the average poverty rate fell by around 12-15 percentage points in crop-based economies (depending on the sample and use of population weights) but hardly changed in resource-based ones.

Table 2.3: $1.25 Poverty in Crop-based vs Resource-based Countries

Panel A	**Crop-based countries**			**Resource-based countries**		
26 countries	**1990**	**2000**	**2010**	**1990**	**2000**	**2010**
Mean	64.6 (63.6)	54.9 (56.4)	43.1 (40.2)	57.9 (52.7)	50.3 (54.9)	48.6 (53.1)
Std. Dev.	19.0 (16.4)	20.1 (18.1)	15.2 (10.8)	19.7 (17.6)	19.7 (19.3)	24.8 (22.1)
Min	18.8	21.2	23.4	24.3	24.4	11.5
Max	93.7	86.4	81.3	83.1	79.4	87.7
N	15	15	15	11	11	11
Pop	186.8	248.6	328.6	195.4	252.5	321.5
% SSA	36.8	37.5	38.0	38.5	38.0	37.2

Panel B	**Crop-based countries**			**Resource-based countries**		
Full sample	**1990**	**2000**	**2010**	**1990**	**2000**	**2010**
Mean	63.5	56.3	39.1**	52.7	56.0	58.0**
Std. Dev.	16.4	17.5	12.6	17.6	19.4	23.5
Min	18.8	21.2	19.8	24.3	24.4	6.1
Max	93.7	86.4	83.7	83.1	80.5	87.7
N	16	19	20	11	13	16
Pop	187.8	266.3	395.7	195.4	266.9	423.9
% SSA	37.0	40.1	45.8	38.5	40.2	49.0

Notes: As for Table 2.2. Panel A is for the 26 countries with data within 5 years either side of 1990, 2000 and 2010; unweighted values listed first with population weighted poverty in parentheses. Panel B is for any country with data for the year, population weighted. The stars on the mean values indicate the level of significance from a difference in means test between crop and resource economies with *** p<0.01, ** p<0.05, * p<0.1.

We explore further why this may have occurred in Table 2.4 where we look at changes in poverty, inequality and output in each of the two decades as well as over the whole 20 year period for both crop and resource-based economies. Changes in poverty can be decomposed into changes in GDP and income distribution, however neither of these two factors appear to sufficiently explain the differing poverty reduction performances

between the two types of economies. One sees that poverty reduction was significantly greater in crop than resource-based economies over the whole period and particularly in the second half, with poverty rates falling by a third from the initial level in crop-based countries. In contrast average poverty, particularly when weighted by population size, in resource-based countries barely changed over the two decades. This divergence in performance was not due to GDP growth as both sets of countries grew at roughly the same rate on average. Nor was it associated with changes in the distribution of income as there is little evidence to suggest there were significant changes on average in the Gini coefficient between the two sets of economies. However, from Table 2.5, we can see that although there were no significant changes to Gini between the two sets, crop-based economies did have significantly lower levels of Gini over the whole period. Thus, given this more equitable distribution of incomes, economic growth was more likely to be poverty reducing in crop-based economies. This may also explain why crop-economies reduced poverty more in the 2000's than over the 1990's, with economic growth almost doubling in the latter period. Although a similar acceleration in economic growth occurred in resource-based economies over the 2000's, as the level of inequality remained high it did not translate to similar falls in poverty levels.

Table 2.6 further separates crop and resource-based countries by their poverty reduction performance. Almost two-thirds of crop-based countries are high performers but this is over 80% in terms of the share of the crop-based population; although a similar share of resource-based countries are also high performers they account for less than a third of the population of resource-based countries. It is also the case that the majority of high performers are crop economies. Most resource-based countries are low performers (especially by population share); although under half the number of low performers, resource countries account for 80% of the total population of low performers. High-performer resource countries are small in terms of population with the exception of South Africa, which accounts for the 51 of the 104 million individuals residing in such countries. Given the problems with South African poverty data (see above) it could be treated as uncertain hence classed as a low performer, in which case the clear majority of people living in resource economies would be in a low performer.

Table 2.4: Changes in poverty, inequality and output in Crop vs Resource Countries

		1990-2000		2000-2010		1990-2010	
	26 countries	**Crop**	**Resource**	**Crop**	**Resource**	**Crop**	**Resource**
Change in Poverty	Mean	-9.7 (-7.3)**	-7.6 (2.0)**	-11.9** (-16.5)***	-1.7** (-2.5)***	-21.5** (-24.0)***	-9.3** (-0.4)***
	Std. Dev.	14.5 (13.2)	13.2 (7.1)	11.8 (12.7)	10.1 (9.4)	16.9 (13.9)	18.1 (11.3)
	Min	-37.4	-35.4	-40.8	-14.7	-53.5	-39.1
	Max	12.3	12.3	5.2	15.1	16.2	20.6
Change in Gini	Mean	-0.6 (-2.4)	-5.2 (-2.2)	-2.2* (-0.2)	1.6* (1.1)	-2.7 (-2.5)	-3.6 (-1.0)
	Std. Dev.	9.6 (8.1)	7.4 (5.1)	4.7 (4.5)	5.5 (3.6)	10.2 (8.3)	5.6 (4.6)
	Min	-12.9	-17.8	-11.8	-4.3	-17.5	-13.4
	Max	22.6	7.2	3.2	12.7	22.4	4.7
Annualised growth in GDP	Mean	2.48 (3.05)	2.81 (2.14)	5.17 (6.17)	4.87 (6.71)	3.80 (4.60)	3.83 (4.39)
	Std. Dev.	2.45 (1.89)	1.45 (1.04)	2.11 (2.29)	2.17 (2.62)	1.73 (1.65)	1.34 (1.43)
	Min	-2.55	1.16	1.10	2.28	0.64	2.14
	Max	6.52	5.45	8.64	8.89	6.95	6.57
N		15	11	15	11	15	11
Pop		248.6	252.5	328.6	321.5	328.6	321.5
% SSA		37.5	38.0	38.0	37.2	38.0	37.2

Note: As for Table 2.3 with unweighted values listed first with population weighted poverty in parentheses. The stars on the mean values indicate the level of significance from a difference in means test between crop and resource economies with *** p<0.01, ** p<0.05, * p<0.1.

Table 2.5: Key Factors in Crop vs Resource Countries

26 countries		Crop-based countries			Resource-based countries		
		1990	**2000**	**2010**	**1990**	**2000**	**2010**
Gini	Mean	41.73*** (40.44)**	41.01*** (38.13)**	38.81*** (37.98)**	56.48*** (49.39)**	51.32*** (47.17)**	52.90*** (48.06)**
	SD	7.59 (6.27)	4.79 (6.19)	5.34 (5.13)	9.06 (8.25)	7.66 (5.96)	7.83 (8.43)
	Min	28.9	30	32.29	42.86	43.25	40.63
	Max	54.14	51.51	51.27	74.33	64.73	64.08
PPE as % of GDP	Mean		6.44* (6.46)**			9.12* (8.00)**	
	SD		1.50 (1.33)			4.05 (2.15)	
	Min		3.89			3.47	
	Max		9.32			16.83	
PPE per capita	Mean		106* (101)			486* (460)	
	SD		57 (59)			538 (473)	
	Min		46			27	
	Max		254			1669	
GDP per capita	Mean	1488** (1332)*	1463** (1333)*	1771** (1709)***	3779** (4007)*	3972** (3760)*	5173** (5264)***
	SD	792 (802)	758 (760)	761 (732)	3214 (3271)	3535 (3155)	4432 (3581)
	Min	653	623	725	462	583	901
	Max	3234	3049	3065	10364	10274	13349
N		15	15	15	11	11	11
Pop		186.8	248.6	328.6	195.4	252.5	321.5
% SSA		36.8	37.5	38.0	38.5	38.0	37.2

Note: As Table 2.4. PPE data is only for 2000 as they are period averages due to data limitations.

We also want to explore the potential impact of expenditure on pro-poor sectors[6] (henceforth PPE) on poverty reduction. Table 2.5 shows that, surprisingly perhaps, PPE as a percentage of GDP and in terms of expenditure per capita is significantly lower in crop-based than resource economies. On average resource-based economies spent a third more on PPE as a proportion of output, but average PPE per capita was over four times larger in resource-based countries given they had far larger GDP per capita. From Table 2.6 we also see that PPE as a percentage of GDP was lowest on average for high performer crop economies. We return to this in the next section which provides statistical tests for the two Africas.

[6] Here defined as expenditure on health and education (see Appendix Table 2.A1); note that Gomanee *et al* (2005) also included expenditure on sanitation (including water).

Table 2.6: $1.25 Poverty Performance and PPE

Panel A (unweighted)	N	2010 Pop	$1.25 1990		$1.25 2010		PPE	
			Mean (SD)	N	Mean (SD)	N	Mean (SD)	N
Crop-based Low Performer	7	62	55.2 (25.5)	5	57.0 (15.4)	6	6.7 (1.2)	7
Resource Low Performer	5	237	57.6 (11.1)	5	64.7 (17.0)	5	8.9 (5.4)	4
Resource High Performer	8	104	58.1 (26.0)	6	39.6 (23.4)	8	8.6 (3.2)	8
Crop-based High Performer	13	297	66.8 (15.7)	11	37.1 (7.8)	12	5.9 (1.6)	13
Total	33	700	61.0 (19.2)	27	46.0 (18.8)	31	7.1 (2.9)	32

Panel B (weighted)	N	2010 Pop	$1.25 1990		$1.25 2010		PPE	
			Mean (SD)	N	Mean (SD)	N	Mean (SD)	N
Crop-based Low Performer	7	62	49.7 (30.0)	5	53.9 (17.3)	6	6.7 (1.4)	7
Resource Low Performer	5	237	56.6 (9.4)	5	61.7 (12.6)	5	7.1 (2.4)	4
Resource High Performer	8	104	42.9 (28.5)	6	35.3 (27.6)	8	8.4 (1.7)	8
Crop-based High Performer	13	297	66.1 (11.6)	11	37.9 (6.0)	12	6.3 (1.4)	13
Total	33	700	58.0 (17.5)	27	47.0 (18.0)	31	6.9 (1.7)	32

Notes: As for Table 2.2; PPE is pro-poor expenditure as a share of GDP (overall average). Panel A not weighted by population, Panel B is. Standard deviations declined over the period (not reported). Sample size varies (no PPE data for Nigeria).

4 Empirical Testing and Validation

Acknowledging the limited quality of the data and given that there are only a few poverty observations for each country, a parsimonious approach is adopted for the econometric analysis: the aim is not to explain variation in poverty or poverty reduction but to test if crop economies do out-perform resource economies. The analysis is based on the World Bank's $1.25 a day measure because it provides the largest sample with consistent measurement. As a robustness check, estimates are also provided for the $1.90 a day measure. One particular problem with poverty analysis is that the poorest people produce only for subsistence and have little or no contact with

the market economy, whereas it is in relation to the subsistence economy that the estimates of national statistical agencies tend to be most inaccurate. This is why many researchers favour the use of household surveys (which do capture subsistence producers and those engaged in informal sectors) over national accounts. Unfortunately, nationally representative household consumption surveys are conducted at best every four or five years and often much less frequently, so even if the sample includes many countries the number of observations (surveys) per country is often quite low. The problem is very much an African one: whereas developing countries have an average of 3.9 estimates of poverty since 2001, African countries have 1.7 (Devarajan, 2013: 59).

In addition to initial poverty and the crop/resource distinction, the core explanatory variables included are growth (a measure of potential to reduce poverty), inequality (to capture political willingness) and pro-poor public expenditure (PPE, to capture policy effort). Although there is acceptance that growth provides the potential for poverty reduction and one expects a significant coefficient, it is clear from Table 2.1 that the elasticity varies significantly. An important factor determining the relationship between growth and poverty reduction is inequality, the distribution of income as measured by the Gini coefficient. That income distribution influences the possibilities for poverty reduction is almost tautologically true, as an increase in the well-being of the poor in relation to the well-being of other income groups automatically makes the distribution of income more equal and reduces the Gini coefficient of inequality. There are plausible reasons why a reduction in inequality might be expected to reduce poverty, including the ability of lower levels of inequality to boost the level of domestic demand and thence production, and the likelihood that lower levels of vertical and horizontal inequality will reduce the likelihood of conflict and thus provide a boost to growth (Mosley, 2012).

Variations in inequality are not an adequate explanation of variations in poverty trajectories. First, there are many cases where poverty goes down over time even though inequality increases as the private 'capitalist sector', in which levels of private enterprise but also inequality are high, expands at the expense of the 'subsistence sector' (there is some evidence for this in Uganda). Furthermore, simply knowing

that inequality matters does not provide us with a pathway explaining how inequality, and thereby poverty, can be reduced in practical terms without prejudicing growth, and such a pathway is what is required by, in particular, those African countries which have not yet been successful in reducing poverty. Fosu (2015) argues that poverty reduction in SSA has been very heterogeneous and seeks to explain that by separating the contribution of growth from that of reductions in inequality for 23 SSA countries and finds that rising inequality contributed in those countries where poverty increased (Guinea-Bissau is the exception where inequality fell) whilst in seven of the 18 countries that reduced poverty changes in inequality limited the reduction. Overall it seems that 'declining inequality tended to decrease poverty (though not necessarily in very low income countries) [while] lower initial inequality raised the rate at which growth was transformed to poverty reduction' (Fosu, 2015: 13).

The important issue may be how rather than whether inequality is reduced, i.e. which groups benefit. As observed in section 2, the key economic reforms in countries such as Ghana and Uganda that reduced poverty were of benefit to agriculture, the sector on which most of the population relied for their livelihood (directly or indirectly). Furthermore, these reforms were initiated by donors and most successfully implemented through poverty reduction strategies in countries that had built a relationship of trust with the aid donors. As donors emphasised the role of social sector spending in poverty reduction, we choose pro-poor public expenditure (PPE) as an indicator of the poverty orientation of policy (Gomanee *et al*, 2005). In the estimation we use per capita PPE rather than as a share of GDP as expenditure in absolute terms better captures the magnitude and extent of poverty reduction policies.

Before discussing the econometric specification (which will explore the differential impacts of growth, inequality, and PPE on poverty between crop and resource economies) it is pertinent to discuss how colonial history may have directly impacted on these factors. Table 2.4 shows that colonial history has not affected the growth rate of economies with both crop and resource growing at similar rates over the 1990s and 2000s. Thus the greater ability of crop economies to reduce poverty has not been as a result of these economies having grown at a faster pace. Similarly, the table also shows that colonial history has not had an impact on the changes in the distribution of income

over the period. Table 2.5 shows that resource economies spent significantly more on PPE (both as a share of GDP and in per capita terms), which indicates that pro-poor policies have not been the direct cause of the divergence in performance between crop and resource economies. However Table 2.5 does reveal one channel through which colonial history may have affected poverty and that is with the level of inequality. One sees that the Gini coefficient both initially and throughout the two decades is significantly lower in crop economies. This is in line with the main hypothesis of this chapter as one would expect crop economies to have a more equitable income distribution given the structure of their colonial economies. Given this greater level of equality, one would expect economic growth to have better translated into increased poverty reduction in crop economies without as great a need for direct pro-poor policy intervention.

As changes in poverty can be mechanically deconstructed into changes in income and the distribution of income, any model of poverty reduction that includes GDP growth and changes in Gini will account for nearly all the variation. Thus such a model should only contain these two independent variables or exclude at least one of them when we are interested in analysing the effect of other variables (such as PPE) on poverty. Consequently the basic specification in general terms is:

$$P_{it} = P(P_0 \text{ , } g \text{ and/or } \Delta Gini, \text{ Crop/Resource, } PPE, [Z]) \qquad (1)$$

Poverty in country *i* in period *t* (the year of a survey observation) is posited to depend on: P_0, initial poverty (the first survey observation in the dataset); *g* (total GDP growth since the initial observation); change in the Gini measure of inequality (since the year of the first observation); Resource (=1 if resource-based and 0 if crop-based) and PPE (in per capita terms averaged over the entire period). A set of additional controls (Z) are added, including fuel exports as a share of GDP, a measure of democracy, and a post-conflict dummy. By including P_0 as a control the regression essentially becomes one looking at changes in poverty since the initial observation. We specify the model in this way rather than having it in first differences (changes in poverty since the previous observation) as we want to capture the impact of colonial history on overall poverty reduction performance and not just over certain periods. In this way a country that greatly reduces poverty over the 1990's for example but less so in the 2000's is

still classified as a successful case of poverty reduction in both periods. However we do also run a first differenced model for robustness, which we will return to later. We also include dummies for the length of time between the current and initial observations to account for the differing time lengths.

The variable of specific interest, *Resource*, acts as an indicator of 'history' (whether the country inherited an inclusive economic structure or otherwise); the controls act as indicators of vulnerability and resilience to shocks (5-year time dummies are also included); while PPE capture ability to deliver pro-poor policy. Pro-poor public expenditure captures the per capita expenditure on health, education and social expenditures (excluding housing), the logic of this being that these are the sectors which have the greatest ability to reduce poverty both by augmenting the human capital of the poor, by providing livelihoods to them and by reducing the cost of the goods which they consume (Gomanee *et al*, 2005). The results presented here are estimated with OLS; Appendix B presents sensitivity analysis with alternative estimators (see below).

The results in Table 2.7 confirm the intuition of Figure 2.1 that the influence of history is critical: *Resource* is a highly significant determinant of poverty and resource-based economies have been much less successful in reducing poverty. Even when *Resource* is insignificant the inference is that given growth and changes in Gini, crop-based economies achieved lower poverty. Initial poverty is clearly important however the suggested baseline of poverty reduction variables between a half and a quarter. Growth is significant and negative; suggesting crop economies tend to have higher growth which associated with lower poverty. The interaction term on growth is positive and although it is not often significant the magnitudes indicate that there is no significant relation between growth and poverty for resource-based economies. Thus one can infer that growth has reduced poverty in crop economies but not in resource. Similarly for changes in Gini and its interaction, the results suggest increases in inequality is associated with higher poverty in crop economies however no such association exists for resource economies. In contrast, higher PPE is associated with lower poverty in resource economies and not crop. A possible inference is that the higher levels of PPE in resource economies permit lower poverty than would otherwise be the case (given

the characteristics of resource economies), whereas PPE is not instrumental in poverty reduction in crop economies.

Table 2.7: Determinants of Poverty Performance (full sample, $1.25 weighted)

	(1)	(2)	(3)	(4)	(5)
P0	0.795***	0.723***	0.668***	0.512***	0.444***
	(0.080)	(0.083)	(0.092)	(0.099)	(0.090)
Resource = 1	6.778**	3.198	-3.970	38.982**	43.693**
	(3.368)	(4.433)	(5.461)	(18.046)	(20.237)
Growth	-0.070**	-0.068*	-0.064*	-0.066**	
	(0.035)	(0.036)	(0.035)	(0.031)	
Resource*G	0.073**	0.039	0.094	0.056	
	(0.032)	(0.057)	(0.064)	(0.060)	
ΔGini			0.571**		0.646***
			(0.231)		(0.213)
Resource*ΔGini			-1.187***		-0.506
			(0.432)		(0.415)
PPE				-1.343	-1.787
				(3.731)	(4.428)
Resource* PPE				-7.049*	-7.324*
				(3.752)	(4.142)
Constant	7.626***	8.199***	13.578***	16.814	24.988
	(0.571)	(0.583)	(2.262)	(19.466)	(23.378)
Observations	97	93	91	93	91
N	33	32	31	32	31
R-squared	0.81	0.81	0.84	0.84	0.86

Note: Estimated by OLS with time dummies for 5 year periods and dummies for the length of time between the current and initial period; dependent variable is period $1.25 a day poverty level weighted by 2010 population; P_0 indicates initial poverty (first observation for the country in the sample); Growth refers to total GDP growth over the period since initial observation; Resource indicates resource-based country; PPE is pro-poor expenditure in per capita terms (overall period average); ΔGini is change in Gini between the current and initial period; N = number countries; Obs = N times periods (surveys). Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. Including PPE loses Nigeria from the sample; (3) reproduces (2) without Nigeria.

In Table 2.7b we rerun the same regressions but with restrictions on the sample and observations so that the initial observations begin at a similar time (from 1988-1993) to reduce the potential bias from time aggregation. In other words we ensure all countries have their initial observation at roughly the same time so that results are not

overly influenced by the differing lengths of time a country is observed for. We find doing so makes little difference to the results. Furthermore very similar results are obtained if the $1.90 poverty line is used (see Table 2.A5).

Table 2.7b: Determinants of Poverty Performance with Additional Time Corrections

	(1)	(2)	(3)	(4)	(5)
P0	0.803***	0.743***	0.709***	0.513***	0.447***
	(0.093)	(0.089)	(0.095)	(0.122)	(0.143)
Resource = 1	-2.549	3.516	-3.243	35.322	43.866
	(4.833)	(7.453)	(7.042)	(21.476)	(28.332)
Growth	-0.141***	-0.140***	-0.105**	-0.087**	
	(0.041)	(0.037)	(0.043)	(0.041)	
Resource*G	0.259***	0.132	0.200	0.208*	
	(0.070)	(0.150)	(0.119)	(0.109)	
ΔGini			0.780**		0.790**
			(0.306)		(0.307)
Resource*ΔGini			-1.559***		-0.958*
			(0.484)		(0.555)
PPE				-3.104	-2.699
				(4.277)	(5.179)
Resource* PPE				-6.648	-6.599
				(4.416)	(5.139)
Constant	7.613	12.326	17.910**	41.820	47.635
	(10.089)	(9.688)	(8.366)	(25.775)	(31.264)
Observations	67	64	62	64	62
N	20	19	18	19	18
R-squared	0.87	0.87	0.91	0.90	0.91

Note: As for Table 2.7 except we limit the sample and observations so that all countries have their initial observation appearing between 1988 and 1993.

Table 2.8: Determinants of $1.25 Poverty (additional variables)

	(1)	(2)	(3)	(4)	(5)	(6)
P0	0.544***	0.401***	0.580***	0.461***	0.404***	0.503***
	(0.099)	(0.085)	(0.109)	(0.093)	(0.104)	(0.103)
Resource = 1	41.735**	68.258***	54.711**	44.755**	59.362***	50.973**
	(17.116)	(20.984)	(21.077)	(20.375)	(22.111)	(23.398)
PPE	-0.645	2.423	1.357	-1.186	-1.365	-0.482
	(3.605)	(3.883)	(4.887)	(4.412)	(4.121)	(5.917)
Resource*	-7.745**	-14.508***	-11.504**	-7.790*	-9.400*	-10.095*
PPE	(3.615)	(4.667)	(4.717)	(4.215)	(4.727)	(5.330)
Growth	-0.077**	-0.111***	-0.069**			
	(0.031)	(0.027)	(0.031)			
Resource* G	0.050	0.057	0.055			
	(0.062)	(0.048)	(0.050)			
ΔGini				0.689***	0.702***	0.664***
				(0.210)	(0.172)	(0.220)
Resource*				-0.613	-0.834**	-0.715**
ΔGini				(0.396)	(0.390)	(0.326)
Post Conf	-7.666			-7.487		
	(5.611)			(4.937)		
Resource*	-4.074			-2.772		
Post Conf	(8.155)			(8.657)		
Agrix		-0.555			0.106	
		(0.351)			(0.407)	
Resource*		-0.345			-1.365*	
Agrix		(0.570)			(0.714)	
Fuelx		-1.407***			-0.940***	
		(0.220)			(0.200)	
Resource*		1.219***			0.603	
Fuelx		(0.345)			(0.400)	
D			-4.544			-0.099
			(5.078)			(5.167)
Resource*D			15.407**			11.850
			(7.140)			(7.615)
Constant	13.002	24.352	-3.824	22.124	28.753	18.029
	(18.828)	(15.591)	(30.369)	(23.086)	(18.903)	(33.814)
Observations	93	92	93	91	90	91
N	32	31	32	31	30	31
R-squared	0.85	0.92	0.85	0.86	0.91	0.87

Note: As for Table 2.7 except: Post Conf is a dummy = 1 for the 10 years following a conflict; Fuelx is fuel and ore exports as % GDP (agriculture and food exports were also included but always insignificant); and D is a binary democracy variable taking the value 1 if the Polity score of democracy (which ranges from 0 to 10) is above 6 and 0 otherwise (all defined in Table 2.A1).

Table 2.8 introduces additional controls into the basic specification. Results for variables in Table 2.7 are effectively unaltered and the additional variables add a little

explanatory power. The post-conflict dummy is negative and significant only for resource economy (once add in interaction term). This is consistent with conflict being a negative shock that increases poverty. The post-conflict period is thus one of poverty reduction (perhaps reflecting a higher starting point as a result of conflict). Variables capturing the composition of exports show interesting results; agricultural exports are significantly associated with lower poverty only in resource economies whereas fuel and ore exports are significantly associated with lower poverty in crop economies. One inference of these results is that the resource curse is binding on resource economies where fuel exports have had no significant impact on reducing poverty but agricultural exports have whereas the opposite is true in crop economies. Democracy, measured as a binary (*D*), has no significant impact on poverty reduction in crop economies but there is slight indication it is associated with higher poverty in resource economies. Similar results are found if the $1.90 poverty line is used except the export share variables generally lose significance for both crop and resource economies (see Table 2.A6); these may be because of the slight change in the sample or because these factors are more important for more extreme levels of poverty (a lower poverty line).

As another way of comparing the two Africas, Table 2.9 reports results separately for crop and resource countries. Interesting growth is now associated with reduced poverty in both crop and resource economies, however as before falling inequality is associated with lower poverty only in crop economies. Similarly, higher PPE is again associated with lower poverty for just resource economies alongside with the post-conflict variables (although this loses significance when using changes in Gini). Agricultural exports have no significant impact for either type of economy but association between higher fuel exports and lower poverty in crop economies is retained. Also retained is the positive correlation between democracy and higher poverty in resource economies. If the $1.90 poverty measure (Table 2.A7) is used then growth is no longer significant in resource economies, but the results for changes in Gini, PPE, and the export share variables are largely the same. Democracy and post-conflict lose significance but one still sees indications they have different impacts on poverty based on the type of economy.

Table 2.9: Determinants of £1.25 Poverty Performance (split sample)

	Crop	Resource	Crop	Resource
P0	0.513***	0.733***	0.444***	0.551
	(0.126)	(0.160)	(0.120)	(0.335)
Growth	-0.101***	-0.127**		
	(0.036)	(0.041)		
ΔGini			0.832***	-0.184
			(0.218)	(0.637)
PPE	0.796	-12.118***	-8.408	-13.882**
	(6.352)	(2.577)	(6.666)	(5.588)
Agrix	-0.173	-0.886	0.488	-0.696
	(0.438)	(0.494)	(0.449)	(1.283)
Fuelx	-1.273***	0.120	-0.687***	0.188
	(0.335)	(0.174)	(0.214)	(0.590)
Post Conf	-0.539	-19.330*	0.362	-18.179
	(5.019)	(10.221)	(4.452)	(12.078)
D	-5.965	18.339***	5.892	21.242*
	(5.910)	(5.454)	(4.922)	(9.671)
Constant	20.943	75.948**	54.021*	85.622
	(25.024)	(32.240)	(29.234)	(50.191)
Observations	57	35	55	35
N	19	12	18	12
R-squared	0.90	0.99	0.93	0.99

Note: As for Table 2.8 except separate crop and resource samples.

Table 2.10 changes the specification used so far slightly by using changes in poverty between each survey as the dependent variable rather than current poverty regressed on initial poverty levels. The main findings regarding the impact of growth and changes in equality in crop and resource economies are preserved. The results for growth are less significant than when using the previous model specification, however the differential impact of growth between the two types of economies are clearly retained. Very similar results are obtained for changes in Gini however PPE is no longer significant for either types of economies. The differential impact of agricultural exports on poverty reduction is even more evident in the new specification, with it increasing poverty in crop and reducing in resource, however fuel exports are no longer significant. These results are more a robustness check to the main specification above as using the short-term changes in poverty will not fully capture the long-run impact of colonial history on overall poverty trajectories. Nevertheless they still imply growth and declining inequality has contributed to poverty reduction in crop economies but not resource.

Table 2.10: Determinants of Poverty Performance using Short-term Changes

	(1)	(2)	(3)	(4)	(5)
Resource = 1	-0.060	1.678	1.408	3.488	0.897
	(1.102)	(4.104)	(4.784)	(3.486)	(4.944)
Growth	-0.230*	-0.246*	-0.217		
	(0.117)	(0.134)	(0.136)		
Resource*Growth	0.181	0.322	0.250		
	(0.230)	(0.271)	(0.275)		
ΔGini	0.509***			0.495**	0.542***
	(0.180)			(0.202)	(0.186)
Resource*ΔGini	-0.245			-0.225	-0.286
	(0.289)			(0.293)	(0.274)
PPE		0.170	-0.813	0.221	-1.182
		(0.699)	(0.934)	(0.660)	(0.870)
Resource*PPE		-0.446	0.239	-0.490	0.689
		(0.782)	(1.047)	(0.738)	(1.007)
Agrix			0.137*		0.169***
			(0.072)		(0.060)
Resource*Agrix			-0.279**		-0.282***
			(0.113)		(0.102)
Fuelx			-0.055		-0.007
			(0.064)		(0.059)
Resource*Fuelx			0.021		-0.030
			(0.090)		(0.087)
Constant	2.271	0.497	2.834	0.504	4.119
	(1.586)	(4.229)	(4.401)	(3.904)	(3.974)
Observations	98	101	100	98	97
N	32	32	31	32	31
R-squared	0.26	0.17	0.23	0.23	0.31

Note: Estimated by OLS weighted by 2010 population with time dummies for 5 year periods; dependent variable is change in the current period $1.25 a day poverty level compared to previous period; Growth refers to total GDP growth over the period between surveys; ΔGini is change in Gini between the current and previous period; the remaining variables are the same as in Tables 2.7 and 2.8. Additional variables from Table 2.8 were tested but always insignificant. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. Nigeria is excluded from the sample in all regressions as including PPE loses Nigeria.

Although the nature of the data limits the estimation strategies that are feasible, Appendix B reports alternatives to OLS: random effects without population weighting and population-averaged estimates using the generalised estimation equation (which allow for the population weights, see Cameron and Trivedi, 2008). The alternative estimators preserve the majority of the main results. In particular the differential

impacts of growth and changes in inequality between crop and resource economies are retained.

5 Conclusion

A political commitment by government is essential to reduce poverty. How can we explain whether an effective pro-poor coalition materialises or not? Several countries that reduced poverty ('improvers' in Table 2.1), such as Sierra Leone, Uganda, Rwanda and Mozambique, may have responded to experience of civil war by realising that policies which favour one regional or ethnic group to the exclusion of others risk re-igniting conflict, and that an inclusive approach to the allocation of government resources is needed to pre-empt that risk. In many cases a broker, who may or may not be an aid donor, is often a vital influence in holding together the members of a pro-poor coalition around an agreed set of pro-poor policies.[7] Finally, to state the obvious, targeting the poor requires pro-poor expenditure programmes and these require resources: the ability to achieve a long-term increase in tax revenue, despite the obvious obstacles to doing this in a fragile state, and to achieve adequate aid inflows to finance a pro-poor development programme may also be important determinants of the ability to put together a lasting coalition (Mosley, 2015),[8] and thence of the poverty trajectory.

The paper argues that colonial experience has been an important factor in laying the foundations for whether the politics that emerged following independence was predisposed to inclusive growth that facilitated poverty reduction. In crop-based colonial economies smallholder producers were economically more important and had

[7] An important case where the key broker was not an aid donor but a civil servant is the case of Uganda in the early 1990s, where the conversion of a reluctant President Museveni to the cause of a pro-poor expenditure programme, supported by the World Bank and other aid donors, was achieved by Emanuel Tumusime-Mutabile, then the Permanent Secretary to the Ministry of Finance. See Mosley (2012: 63-64)

[8] Aid donors supply not only money, of course, but also policy advice. The impact of policy conditionality is highly controversial, but Gomanee *et al* (2005) suggest that during the 1990s and early 2000s conditionality was an important influence on developing countries adopting pro-poor expenditure programmes, especially in the health and education fields.

potentially greater political weight after independence (although this was not always realised). Resource-based colonial economies, in contrast, fostered a less inclusive politics after independence (subsequent major resource discoveries had a similar effect). This, combined with the tendency of resource sectors to have relatively weaker linkages with the rest of economy, implied that growth was less likely to translate into poverty reduction. The major finding of the analysis in the paper is that crop economies performed significantly better in reducing poverty (since around 1990 when growth took off). On average, crop economies reduced poverty by over a third in the two decades after 1990 whereas resource economies, on average, failed to reduce poverty significantly. Growth in GDP and having higher resource exports are the principal determinants of the rate of poverty reduction in crop economies, whereas neither are significant in resource economies. The main finding is that crop economies are better able than resource economies to translate growth into poverty reduction.

The essence of our story consists of three elements. First, *history*: colonial and post-colonial policies determine economic structure (in particular a country's status as a 'mine-plantation type' or a 'peasant export-type' economy), the extent of linkages from the core sector to the economy (through demand for factors and inputs and supply of inputs to further processing) and the initial distribution of economic opportunities between rich and poor. Second, *shocks* (and defences against them) affect the historical legacy by either consolidating determined government policies and institutional reforms which bring a lasting pro-poor political settlement, or eroding them through the effect of lasting shocks and malfunctioning institutions. In this context resource discoveries are an important shock. Third, *policies* are fundamental as lasting pro-poor political settlements will only happen if governments are motivated to make them happen, in other words if a pro-poor political economy can be brought into being and embodied in a pro-poor ruling coalition.

In contrast to a literature which has tended to analyse African poverty trends in an aggregative way as either good or bad for the continent as a whole, the focus here is on inter-country differences, which are dramatic. There appear to be two distinct groups of SSA countries, in one of which the gains from growth have been widely spread among the population whereas in the other only a select minority have been

able to climb the income ladder, and colonial legacy is an important factor in distinguishing the two groups. Sadly, the Africa which provides the most illustrations of dramatic improvement in the living standards of the poor consists mainly of small countries: with the exception of Ethiopia, no large country in Africa has cracked the problem of how to cut poverty fast. There is a very wide variance between the best and the worst cases, a phenomenon which has been little discussed, and our purpose has been to understand this variance.

The more inclusive of the two Africas, in most but not all cases, consists of countries which in colonial times relied on African smallholder exports to achieve viability – thereby providing a potential vehicle for broad-based development, as a high proportion of low-income Africans nonetheless have a little tea, coffee, cotton or cocoa on their plots and many more work for people who do. But not all former 'peasant export economies', by any means, have been able to capitalise on this asset, nor have the countries which lacked it been barred from overcoming the barriers to successful poverty reduction – provided they had the necessary political will, and the necessary capacity and imagination to assemble a set of policy instruments appropriate to local circumstances. The paper explores the nature of that set of policy instruments and what might be the secret of forming the 'pro-poor coalitions' required to convert policy instruments into a workable development strategy.

The analysis focused on public expenditure not because other policy tools are not important, but because without investment in human capital (through social sector expenditures) most of the other instruments will not work. There is some evidence that a pro-poor expenditure pattern influences poverty reduction in resource economies (where the level is higher than in crop economies) – redistributive spending can counteract the other characteristics that act against poverty reduction. In crop economies, in contrast, higher pro-poor spending is associated with higher poverty (perhaps because, ceteris paribus, higher poverty motivates spending). The main motivational factors bringing into being a pro-poor political settlement, and thus a pro-poor expenditure pattern, appear to be above all an adequate capacity to generate revenue, and in certain cases recent experience of conflict, which appears to 'concentrate minds' and induce a more long-term and inclusive approach to policy. A

broker able to act as an intermediary between aid donors and government also appears to be an important factor influencing the pattern of public spending.

These lessons must be stated with caution, partly because the statistics are questionable and partly because so many of the influences which determine poverty appear to be local. Maybe the most crucial influence not yet properly examined is the political significance of smallholder agricultural interests. This would be difficult to quantify (the crop base is the proxy used here), but where smallholders are politically strong, poverty reduction performance has been good, and wherever they are weak poverty reduction performance has been bad. Trying to convert this intuition into a serious investigation of the mechanisms through which it works (if it does work) represents an important frontier for future research.

Appendix to Chapter 2

Appendix A

Table 2.A1: Data Sources and Definitions

Variable	Indicator	Source
GDP per capita	GDP per capita	World Development Indicators 2012 ("WDI 2012"). Available online: http://data.worldbank.org/data-catalog/world-development-indicators
Poverty $1.25	$1.25 and $2 per day (2005 PPP); National poverty line	World Bank POVCAL data from http://iresearch.worldbank.org/PovcalNet/index.htm?2, accessed 23rd April 2015
Poverty $1.90	$1.90 per day (2010 PPP)	World Bank POVCAL data from http://iresearch.worldbank.org/PovcalNet/index.htm?2, accessed 3rd November 2015
Colonial heritage	Crop-based (Cash crop) Resource-based (Labour Reserve and Concession Companies)	Adapted from Mkandawire (2010)
Growth	Average GDP growth rate between poverty observations	World Development Indicators 2012 ("WDI 2012"). Available online: http://data.worldbank.org/data-catalog/world-development-indicators
Gini	Gini coefficient	World Bank POVCAL data from http://iresearch.worldbank.org/PovcalNet/index.htm?2, accessed 23rd April 2015
PPE	Public expenditure on health and education in per capita terms averaged over years for which data are available from 1980-2013.	Calculated using data from World Development Indicators 2012 ("WDI 2012"). Available online: http://data.worldbank.org/data-catalog/world-development-indicators
Democracy	1= democracy 0 =non democratic	Based on Polity II Dataset. Values above 6 are considered democracies http://www.systemicpeace.org/polity/polity4.htm

Polity	Polity measure is a raw score ranging from -10 to 10	Polity II Dataset. http://www.systemicpeace.org/polity/polity4.htm
Post Conflict	1= Within 10 years after a conflict 0 = Currently in conflict or >10 since a conflict	Based on data from Uppsala Conflict Data Program (UCDP). http://www.pcr.uu.se/research/ucdp/datasets/
Agrix	Food and agricultural raw materials exports as % of GDP averaged over years data is available from 1980-2013.	Calculated using data from World Development Indicators 2012 ("WDI 2012"). Available online: http://data.worldbank.org/data-catalog/world-development-indicators
Fuelx	Fuel and ore exports as % of GDP averaged over years data is available from 1980-2013.	Calculated using data from World Development Indicators 2012 ("WDI 2012"). Available online: http://data.worldbank.org/data-catalog/world-development-indicators

Note: Data file available on request.

List of countries coded as democracies (D=1, otherwise 0)

Benin (2006-2010), Botswana (1987-2010), Gambia (1980-1993), Ghana (2004-2010), Kenya (2002-2010), Madagascar (1992-2008), Mali (1992- 1996; 2002-2010), Niger (1992-1995), Nigeria (1980-1983), Senegal (2000-2010), Sierra Leone (2007-2010), South Africa (1993-2010), Zambia (2008-2010).

Table 2.A2: Sample Countries and $1.90 Poverty Performance 1990-2013

Country	Period	Sur-veys	% Δ GDP pa	$1.90 Poverty Headcount			Elast-icity
				Initial	Current	% Δ pa	
Gambia	1998-2003	2	4.1	73.8	45.3	-9.3	-2.2
Congo, Rep.	2005-2011	2	5.4	50.2	28.7	-8.9	-1.7
Mauritania	1987-2008	6	1.4	40.1	10.9	-6.0	-4.3
Chad	2003-2011	2	10.1	62.9	38.4	-6.0	-0.6
Namibia	1994-2010	3	4.2	52.9	22.6	-5.2	-1.2
Ghana	1988-2006	5	4.7	63.8	25.2	-5.0	-1.1
Swaziland	1995-2009	3	2.7	81.7	42.0	-4.6	-1.7
Guinea	1991-2012	5	3.3	92.3	35.3	-4.5	-1.3
Cameroon	1996-2007	3	4.0	48.1	29.3	-4.4	-1.1
Ethiopia	1995-2011	4	7.5	67.9	33.5	-4.3	-0.6
Uganda	1989-2012	8	6.8	88.0	33.2	-4.1	-0.6
Botswana	1986-2009	4	5.8	42.6	18.2	-3.6	-0.6
South Africa	1993-2011	6	3.2	31.9	16.6	-3.6	-1.1
Mali	1994-2010	4	5.5	84.9	49.3	-3.3	-0.6
Senegal	1991-2011	5	3.5	68.0	38.0	-2.9	-0.8
Burkina Faso	1994-2009	4	5.8	83.1	55.3	-2.7	-0.5
Niger	1993-2011	5	3.7	78.2	50.3	-2.4	-0.6
Congo, Dem Rep	2005-2012	2	6.0	91.2	77.2	-2.4	-0.4
Tanzania	1992-2012	4	5.3	70.4	46.6	-2.0	-0.4
Mozambique	1996-2009	3	7.7	85.4	68.7	-1.7	-0.2
C. African Rep.	1992-2008	3	2.5	84.3	66.3	-1.5	-0.6
Sierra Leone	1990-2011	3	2.4	65.5	52.3	-1.1	-0.4
Lesotho	1987-2010	4	4.1	74.8	59.7	-1.0	-0.2
Weighted Average 'High Performers'			**5.6**	**70.1**	**42.6**	**-3.5**	**-0.7**
Togo	2006-2011	2	3.5	55.6	54.2	-0.5	-0.1
Burundi	1992-2006	3	-0.5	81.1	77.7	-0.3	0.6
Rwanda	1985-2011	4	4.8	63.7	60.3	-0.2	0.0
Malawi	1998-2010	3	4.1	63.6	70.9	0.9	0.2
Benin	2003-2011	2	3.5	48.9	53.1	1.1	0.3
Weighted Average 'Uncertain'			**3.2**	**63.1**	**64.5**	**0.3**	**0.2**
Nigeria	1986-2010	5	4.7	46.0	53.5	0.6	0.1
Zambia	1991-2010	8	4.6	54.1	64.4	0.9	0.2
Madagascar	1993-2010	6	2.8	69.4	81.8	1.0	0.3
Guinea-Bissau	1991-2010	4	1.9	43.0	67.1	2.4	1.3
Kenya	1992-2005	4	2.5	23.1	33.6	2.9	1.2
Cote d'Ivoire	1985-2008	10	1.7	6.8	29.0	6.5	3.9
Weighted Average 'Deteriorating'			**3.9**	**41.8**	**51.5**	**1.5**	**0.6**

Notes: As for Table 2.1.

Table 2.A3: Tests for Difference in Group Means ($1.25)

Panel A: P-values for T-tests on unweighted means

Comparison	% Δ GDP pa	$1.25 Poverty Headcount			Elasticity
		Initial	Current	% Δ pa	
Improvers vs Uncertain	0.039	0.328	0.001	0.000	0.000
Improvers vs Deteriorating	0.015	0.040	0.026	0.015	0.054
Improvers vs Low Reducers	0.004	0.055	0.000	0.001	0.016
Uncertain vs Deteriorating	0.462	0.072	0.432	0.090	0.118

Panel B: Unweighted mean differences

Comparison	% Δ GDP pa	$1.25 Poverty Headcount			Elasticity
		Initial	Current	% Δ pa	
Improvers vs Uncertain	1.8	3.0	-22.5	-3.1	-0.8
Improvers vs Deteriorating	1.9	19.3	-20.8	-6.6	-2.7
Improvers vs Low Reducers	1.8	11.2	-21.6	-4.9	-1.8
Uncertain vs Deteriorating	0.1	16.3	1.7	-3.5	-1.9

Notes: Tests for statistical differences in means for $1.25 a day poverty reported in Table 2.1. Low Reducers are Uncertain plus Deteriorating groups. Panel B reports differences in means, Panel A reports p-values for T-tests on means.

Table 2.A4: Tests for Difference in Group Means ($1.90)

Panel A: P-values for t-tests on Unweighted means

Comparison	% Δ GDP pa	$1.90 Poverty Headcount			Elasticity
		Initial	Current	% Δ pa	
Improvers vs Uncertain	0.075	0.184	0.002	0.000	0.000
Improvers vs Deteriorating	0.013	0.012	0.085	0.000	0.006
Improvers vs Low Reducers	0.007	0.011	0.005	0.000	0.000
Uncertain vs Deteriorating	0.483	0.035	0.206	0.031	0.081

Panel B: Unweighted mean differences

Comparison	% Δ GDP pa	$1.90 Poverty Headcount			Elasticity
		Initial	Current	% Δ pa	
Improvers vs Uncertain	1.7	6.2	-22.2	-4.1	-1.2
Improvers vs Deteriorating	1.7	28.4	-13.9	-6.3	-2.2
Improvers vs Low Reducers	1.7	18.3	-17.7	-5.3	-1.7
Uncertain vs Deteriorating	0.0	22.2	8.3	-2.2	-1.0

Notes: As for Table 2.A3 except refers to $1.90 a day poverty reported in Table 2.A2.

Table 2.A5: Determinants of Poverty Performance ($1.90 headcount)

	(1)	(2)	(3)	(4)	(5)
P0	0.757***	0.743***	0.688***	0.599***	0.381***
	(0.084)	(0.071)	(0.070)	(0.066)	(0.087)
Resource = 1	6.771	-1.716	-9.283*	32.709**	47.706**
	(5.780)	(4.671)	(4.964)	(14.126)	(20.382)
Growth	-0.134***	-0.147***	-0.143***	-0.135***	
	(0.037)	(0.030)	(0.025)	(0.023)	
Resource*G	0.096	0.081	0.140***	0.095**	
	(0.060)	(0.053)	(0.049)	(0.039)	
ΔGini			0.580***		0.632***
			(0.213)		(0.207)
Resource*ΔGini			-1.395***		-0.305
			(0.398)		(0.393)
PPE				-1.508	-1.640
				(3.307)	(4.877)
Resource* PPE				-6.483**	-8.339*
				(3.208)	(4.654)
Constant	6.070***	6.221***	11.612***	14.917	22.021
	(0.515)	(0.431)	(1.955)	(17.203)	(25.631)
Observations	99	95	91	95	91
N	34	33	31	33	31
R-squared	0.76	0.84	0.86	0.88	0.84

Note: Corresponds to Table 2.7 except for $1.90 a day poverty measure.

Table 2.A6: Determinants of $1.90 Poverty (additional variables)

	(1)	(2)	(3)	(4)	(5)	(6)
P0	0.606***	0.510***	0.629***	0.383***	0.327***	0.390***
	(0.071)	(0.071)	(0.069)	(0.084)	(0.105)	(0.091)
Resource = 1	34.248**	41.538**	42.002***	48.179**	54.071**	51.251**
	(14.389)	(19.248)	(15.320)	(20.556)	(26.405)	(22.505)
PPE	-1.499	1.653	1.260	-1.907	-0.031	-0.477
	(3.389)	(4.254)	(3.880)	(5.045)	(6.498)	(5.644)
Resource*	-6.707**	-10.814**	-8.648**	-8.236*	-10.685	-9.007*
PPE	(3.273)	(4.572)	(3.684)	(4.703)	(7.044)	(5.196)
Growth	-0.139***	-0.147***	-0.142***			
	(0.028)	(0.024)	(0.020)			
Resource* G	0.095**	0.121***	0.083**			
	(0.040)	(0.038)	(0.037)			
ΔGini				0.623***	0.628***	0.577**
				(0.209)	(0.212)	(0.250)
Resource*				-0.311	-0.269	-0.243
ΔGini				(0.398)	(0.406)	(0.393)
Post Conf	-0.366			4.672		
	(5.082)			(5.313)		
Resource*	-11.738**			-10.404		
Post Conf	(5.527)			(7.327)		
Agrix		-0.548			-0.306	
		(0.406)			(0.616)	
Resource*		0.682			0.351	
Agrix		(0.598)			(0.783)	
Fuelx		-0.809***			-0.318	
		(0.268)			(0.353)	
Resource*		0.655*			0.154	
Fuelx		(0.386)			(0.481)	
D			-10.134**			-3.339
			(4.596)			(7.196)
Resource*D			8.629			1.241
			(6.119)			(7.933)
Constant	14.845	19.982	0.480	23.309	24.500	15.488
	(17.636)	(17.184)	(20.197)	(26.548)	(25.979)	(30.226)
Observations	95	93	95	91	90	91
N	33	32	33	31	30	31
R-squared	0.88	0.90	0.89	0.84	0.85	0.84

Note: Corresponds to Table 2.8 except for $1.90 a day poverty measure.

Table 2.A7: Determinants of $1.90 Poverty Performance (split sample)

	Crop	Resource	Crop	Resource
P0	0.520***	0.481**	0.287*	0.438*
	(0.137)	(0.157)	(0.154)	(0.210)
Growth	-0.153***	-0.053		
	(0.031)	(0.054)		
ΔGini			0.587*	-0.052
			(0.289)	(0.469)
PPE	3.081	-9.906***	-2.691	-10.071**
	(4.662)	(2.638)	(6.222)	(4.282)
Agrix	-0.188	-0.385	0.075	-0.335
	(0.553)	(0.796)	(0.861)	(0.862)
Fuelx	-0.967***	-0.207	-0.404	-0.192
	(0.270)	(0.415)	(0.349)	(0.547)
Post Conf	3.167	-12.179	10.188*	-11.238
	(5.006)	(17.330)	(5.531)	(15.837)
D	-12.859**	-2.497	-3.501	-0.629
	(5.512)	(10.142)	(8.631)	(10.064)
Constant	4.745	104.284**	29.149	102.331*
	(24.411)	(40.147)	(29.600)	(47.691)
Observations	59	34	56	34
N	19	12	18	12
R-squared	0.90	0.98	0.86	0.97

Note: Corresponds to Table 2.9 except for $1.90 a day poverty measure.

Appendix B

Replications with Alternative Estimators

The first robustness check is to use a random effects estimator (fixed effects is not feasible given the number of time-invariant variables); Breusch and Pagan Lagrangian multiplier tests support the use of random effects for all cases. This allows for the inclusion of individual country effects whilst imposing the assumption that these effects are not correlated with the independent variables. Tables 2.B1-2.B3 replicate the results of Tables 2.7, 2.8 and 2.9 in the text using random effects. The results are very similar for the main variables of interest (Resource, Growth and ΔGini); albeit with Resource having smaller coefficients and as a result being less significant or losing insignificance. Growth and ΔGini as well as their interactions with Resource are generally the same in terms of size and significance apart from a few cases where they are less significant. Although the interaction term for PPE is less significant when using Random Effects, when the combined PPE effect is calculated for resource economies the size and significance level of the OLS regressions are retained. Significance levels are reduced for some of the additional variables (particularly the democracy dummy) but these are not major differences (Table 2.B2). When the sample is split (Table 2.B3), little is changed apart from growth no longer being significant for resource economies.

As one can no longer do population weighting with random effects, Tables 2.B4-2.B6 replicate the results of Tables 2.7-2.9 in the text using a population-averaged estimation method which allows for population weighting. The population-averaged estimates are obtained using the generalised estimation equation (GEE) approach whereby specifying a normal distribution for the dependent variable, using an identity link function and assuming equicorrelated/exchangeable errors gives estimates which are asymptotically equivalent to random effects estimation (Cameron and Trivedi, 2008).

The results are again very similar for the main variables and if anything provide even stronger evidence for many of the patterns we found using OLS. The resource dummy is generally slightly larger in size but of the same significance level in all cases.

Growth and its interaction (particularly the latter) are either of a higher significance level or become significant when they were insignificant before. This is also true for PPE and its interaction and to some extend for ΔGini and its interaction. Significance is also mostly retained for the additional variables in Table 2.B5. When the sample is split (Table 2.B6); again results are largely the same apart for some variables becoming significant whereas before they were not. These changes conform with the general patterns highlighted before, for example Agricultural exports is now shown to significantly reduced poverty in resource countries but significantly increases it in crop.

The Pooled OLS estimates in Tables 2.7-2.10 all use robust standard errors which accounts for heteroskedasticity. Clustering the robust standard errors at the country level would also allow for serial correlation. However given there are quite large gaps between the time observations for each country and that many of our variables of interest are fixed over time, serial correlation is unlikely to be big problem. Furthermore clustering is recommended when T is large and as $N \rightarrow \infty$ (Cameron and Trivedi, 2008: 832); thus as T in our data is generally very small and N of a moderate size we have left the robust standard errors unclustered.

In general the alternative estimators preserve the main results and patterns highlighted in the main analysis. Results for the split sample appear to be the most robust and do suggest that many variables, particularly Gini and the composition of exports, have significantly different impacts on crop and resource economies.

Table 2.B1: Replication of Table 2.7 with Random Effects Estimation

	(1)	(2)	(3)	(4)	(5)
P0	0.624***	0.614***	0.651***	0.385***	0.346***
	(0.138)	(0.140)	(0.142)	(0.136)	(0.103)
Resource = 1	-2.721	-5.581	5.305	28.880	48.819
	(5.433)	(5.733)	(8.271)	(33.134)	(30.588)
Growth	-0.121***	-0.129***	-0.087*	-0.079	
	(0.038)	(0.040)	(0.049)	(0.050)	
Resource*G	0.072*	0.083	-0.025	0.075	
	(0.037)	(0.054)	(0.083)	(0.059)	
ΔGini			0.808***		0.838***
			(0.223)		(0.197)
Resource*ΔGini			-0.727		-0.389
			(0.609)		(0.336)
PPE				-7.229	-4.674
				(7.225)	(6.680)
Resource* PPE				-4.296	-7.559
				(7.052)	(6.462)
Constant	10.923		28.220		
	(14.792)		(20.606)		
Observations	97	93	91	93	91
N	33	32	31	32	31

Note: As for Table 2.7 but unweighted and using Random Effects Estimation.

Table 2.B2: Replication of Table 2.8 with Random Effects Estimation

	(1)	(2)	(3)	(4)	(5)	(6)
P0	0.386***	0.362***	0.451***	0.342***	0.327***	0.373***
	(0.133)	(0.101)	(0.146)	(0.102)	(0.095)	(0.114)
Resource = 1	35.808	60.159*	40.027	55.461*	54.867**	45.245
	(32.057)	(31.125)	(33.026)	(30.512)	(24.208)	(30.073)
PPE	-6.725	0.069	-4.816	-4.226	-3.559	-4.687
	(7.275)	(6.349)	(7.869)	(6.654)	(4.573)	(7.424)
Resource*	-5.618	-11.800*	-7.702	-8.771	-8.496*	-8.240
PPE	(7.044)	(6.590)	(7.400)	(6.519)	(4.820)	(6.913)
Growth	-0.086*	-0.103***	-0.083			
	(0.050)	(0.035)	(0.053)			
Resource* G	0.082	0.053	0.062			
	(0.065)	(0.044)	(0.055)			
ΔGini				0.874***	0.799***	0.888***
				(0.197)	(0.153)	(0.177)
Resource*				-0.498	-0.715*	-0.864***
ΔGini				(0.323)	(0.392)	(0.327)
Post Conf	-2.979			-2.361		
	(5.845)			(5.015)		
Resource*	-8.452			-9.962		
Post Conf	(10.656)			(9.976)		
Agrix		-0.344			0.174	
		(0.430)			(0.380)	
Resource*		-0.464			-1.051	
Agrix		(0.692)			(0.769)	
Fuelx		-1.143***			-0.866***	
		(0.205)			(0.142)	
Resource*		0.841**			0.510	
Fuelx		(0.378)			(0.417)	
D			1.106			5.735
			(7.191)			(4.809)
Resource*D			11.541			6.552
			(8.812)			(6.461)
Observations	93	92	93	91	90	91
N	32	31	32	31	30	31

Note: As for Table 2.8 but unweighted and using Random Effects Estimation.

Table 2.B3: Replication of Table 2.9 with Random Effects Estimation

	Crop	Resource	Crop	Resource
P0	0.399**	0.590***	0.394***	0.508***
	(0.184)	(0.100)	(0.118)	(0.183)
Growth	-0.108**	-0.094		
	(0.048)	(0.068)		
ΔGini			0.906***	-0.280
			(0.144)	(0.488)
PPE	1.459	-12.900***	-6.928	-14.041***
	(6.379)	(1.529)	(6.362)	(2.903)
Agrix	-0.214	-0.645	0.515	-0.713
	(0.537)	(0.677)	(0.441)	(0.887)
Fuelx	-1.108***	0.218	-0.598***	0.133
	(0.245)	(0.394)	(0.157)	(0.520)
Post Conf	-1.342	-21.055	0.204	-23.227*
	(5.474)	(13.890)	(4.275)	(13.231)
D	-9.525	15.953**	2.321	16.730**
	(6.220)	(7.689)	(5.026)	(7.124)
Observations	57	35	55	35
N	19	12	18	12

Note: As for Table 2.9 but unweighted and using Random Effects Estimation.

Table 2.B4: Replication of Table 2.7 with Population Averaged Model

	(1)	(2)	(3)	(4)	(5)
P0	0.790***	0.689***	0.746***	0.557***	0.492***
	(0.119)	(0.110)	(0.109)	(0.113)	(0.088)
Resource = 1	7.067*	-4.827	-6.886	50.257**	48.623**
	(3.909)	(3.894)	(4.322)	(22.921)	(22.414)
Growth	-0.092***	-0.095***	-0.115***	-0.097***	
	(0.027)	(0.028)	(0.027)	(0.028)	
Resource*G	0.059***	0.114***	0.110**	0.107**	
	(0.021)	(0.043)	(0.053)	(0.044)	
ΔGini			0.566***		0.555***
			(0.158)		(0.179)
Resource*ΔGini			-0.824***		-0.403
			(0.271)		(0.288)
PPE				3.988	0.795
				(5.124)	(4.864)
Resource* PPE				-11.194**	-8.867*
				(5.138)	(4.817)
Observations	97	93	87	93	91
N	33	32	31	32	31

Note: As for Table 2.7 but using Population Averaged Model.

Table 2.B5: Replication of Table 2.8 with Population Averaged Model

	(1)	(2)	(3)	(4)	(5)	(6)
P0	0.559***	0.415***	0.590***	0.496***	0.385***	0.543***
	(0.112)	(0.071)	(0.109)	(0.084)	(0.051)	(0.096)
Resource = 1	54.020**	69.752***	52.986**	48.036**	61.525***	53.413**
	(23.241)	(24.190)	(21.946)	(22.037)	(14.968)	(22.338)
PPE	4.369	3.578	3.670	0.568	-1.656	1.148
	(5.145)	(4.016)	(5.422)	(4.555)	(2.615)	(5.758)
Resource*	-11.870**	-15.141***	-12.016**	-8.808*	-9.598***	-10.788**
PPE	(5.224)	(5.234)	(4.995)	(4.702)	(3.128)	(5.144)
Growth	-0.098***	-0.118***	-0.097***			
	(0.029)	(0.023)	(0.028)			
Resource* G	0.107**	0.078**	0.097**			
	(0.044)	(0.034)	(0.041)			
ΔGini				0.623***	0.717***	0.588***
				(0.171)	(0.109)	(0.178)
Resource*				-0.501*	-0.940***	-0.570**
ΔGini				(0.283)	(0.287)	(0.262)
Post Conf	-0.072			-5.135		
	(2.532)			(3.364)		
Resource*	-11.380*			-4.507		
Post Conf	(6.743)			(7.197)		
Agrix		-0.468			0.062	
		(0.361)			(0.211)	
Resource*		-0.343			-1.510***	
Agrix		(0.545)			(0.396)	
Fuelx		-1.423***			-0.885***	
		(0.226)			(0.091)	
Resource*		1.149***			0.611	
Fuelx		(0.327)			(0.222)	
D			0.253			0.289
			(2.786)			(3.610)
Resource*D			6.067			10.447**
			(4.295)			(4.940)
Observations	93	92	93	91	90	91
N	32	31	32	31	30	31

Note: As for Table 2.8 but using Population Averaged Model.

Table 2.B6: Replication of Table 2.9 with Population Averaged Model

	Crop	Resource	Crop	Resource
P0	0.486***	0.724***	0.432***	0.558***
	(0.097)	(0.070)	(0.056)	(0.147)
Growth	-0.094***	-0.124***		
	(0.020)	(0.021)		
ΔGini			0.856***	-0.223
			(0.112)	(0.359)
PPE	-1.470	-12.350***	-7.586**	-14.033***
	(4.103)	(0.981)	(3.823)	(2.275)
Agrix	-0.080	-0.876***	0.452**	-0.792
	(0.347)	(0.208)	(0.209)	(0.725)
Fuelx	-1.233***	0.146**	-0.630***	0.196
	(0.254)	(0.071)	(0.110)	(0.344)
Post Conf	1.211	-20.182***	-0.314	-19.414***
	(3.640)	(5.909)	(2.855)	(6.339)
D	-2.102	18.905***	5.472**	22.216***
	(3.775)	(3.387)	(2.472)	(5.910)
Observations	57	35	55	35
N	19	12	18	12

Note: As for Table 2.9 but using Population Averaged Model.

Chapter 3:

Constructing Cohorts to Obtain Consistent Pseudo-Panel Estimates

Abstract

Pseudo-panels enable the estimation of panel models when only repeated cross-sections rather than panel data are available. It involves the grouping of individual into cohorts and using the cohort means as if they are observations in a genuine panel. The usual assertion is that as long as there are enough individuals within cohorts so that the cohort sample means are a good approximation of the cohort population means then pseudo-panel estimates are consistent, otherwise they may suffer from measurement error bias. We show that there can be substantial bias arising from the grouping process itself due to the loss of variation and heterogeneity as one moves from the individual to the cohort level. Thus we find many of the common methods used for grouping into cohorts produces inconsistent estimates. We develop some measures for assessing whether the cohort level data contains sufficient variation for pseudo-panel estimates to be consistent, focusing on the variation across cohort means, over time and within cohort groups. We then test the measures empirically and with Monte Carlo simulations, providing useful thresholds that can be used to accept or reject the cohort construction method.

1 Introduction

The advantages of using panel datasets, which include both time-series and cross-section dimensions, for empirical analysis are well known. However, in many settings such data may not be available due to the cost and difficultly of following the same set of individual agents over a sufficiently long period of time. Instead what often is available is repeated cross-sections (henceforth RCS) where a different set of individuals are observed in each time period. This is particularly true for household surveys both in developed and developing countries. The obvious disadvantage of RCS data are that many estimation techniques available to panel datasets, such as controlling for fixed effects, estimating a differenced model or a dynamic model is not possible.

The seminal work of Deaton (1985) proposed a methodology for the consistent estimation of fixed effect models using RCS data. The basic idea is to construct cohorts and track their progress over time. A cohort is defined as a group of individuals who share a common characteristic which is constant over time and can be traced over the different survey years, one oft used example is an age cohort such as all individuals born between 1945 and 1950. Then the sample means of each cohort in each time period are calculated and treated as if they are observations in a genuine panel, called a pseudo or synthetic panel, but subject to measurement error. Applications of pseudo-panel estimation initially started in the consumption and labour supply literature with Browning, Deaton, and Irish (1985) but has since been increasingly used in various fields ranging from estimating the private demand for medical insurance in Britain (Propper, Rees and Green, 2001) to agricultural production (Heshmati and Kumbhakar, 1997) and (Paul and Nehring, 2005). Applications have also extended beyond the simple linear model to linear dynamic models (Dargay and Vythoulkas, 1999), nonlinear dynamic models (Antman and McKenzie, 2007b) and duration analysis (Güell and Hu, 2006). One particular field where pseudo-panel techniques have started to become popular is development economics; some examples include Nicita (2009), Warunsiri and McNown (2010) and Dang et al. (2014). This trend is likely to continue as the World Bank since the

mid-1980's has been helping many countries produce repeated surveys detailing household behaviour under its Living Standards Measurement Survey (LSMS), which now contains data for about 40 different countries. Thus pseudo-panel estimation techniques has the potential to become an important tool in the analysis of developing country household decisions. The question naturally arises on how reliable pseudo-panel estimates are particularly when used in this context.

Although a literature has developed on the conditions for the consistency of pseudo-panel estimates the main focus has been on how to deal with bias resulting from measurement error and whether there is a sufficient number of individuals within a cohort for the cohort sample means to be representative of the population means. While cognizant of these issues, we focus on the practicalities of cohort construction in order to obtain some guidelines of how it should be conducted as this has gone largely unaddressed in the literature. We find that even when the measurement error bias is minimised, by using panel data, pseudo-panel estimates are still highly sensitive to the way cohorts are constructed and, contrary to expectations, the best estimates are derived when the number of individuals within a cohort is lowest. We argue these results are due to bias arising from the aggregation process which leads to loss of variation in the cohort level data. Furthermore ignoring such bias results in many of the methods commonly used for cohort construction being unsuitable, particularly if one is interested in household or individual level analysis. However if sufficient variation exists at the cohort level we find that pseudo-panel estimates can mimic results obtained when using genuine panel estimation.

We first try to identify whether certain variables with which cohorts can be constructed are better than others at producing accurate estimates. We fail to draw any general conclusions other than then recommending the use of age cohorts with year bands of 5 years or less. We then derive some measures, which we call AWAR and CAWAR, for assessing the variation created at the cohort level, identifying some useful thresholds for accepting or rejecting whether pseudo-panel estimates are likely to be reliable. This is done using both an empirical application and results from Monte Carlo simulations.

Section 2 expands on the potential sources of bias arising from the cohort construction process and motivates our interest in focusing on the loss of variation during aggregation. Section 3 details the data and methodology used in order to assess the impact of cohort construction on estimates while minimising the effect of measurement error bias. Section 4 estimates the pseudo-panels based on an exhaustive approach to cohort construction, demonstrating the presence of aggregation bias, and highlighting the bias in estimates based on popular cohort construction methods. In section 5 we analyse the relative merits of different variables which can be used to construct cohorts. Section 6 derives the AWAR measure for assessing whether the cohort level data has sufficient variation in order to deliver consistent pseudo-panel estimates as well as giving an example of its application. We also show that cohort construction is similar to choosing an instrumental variable and use some existing tests of instrument validity to help assess cohort aggregation and compare them to AWAR. Section 7 conducts some Monte Carlo simulations to test the generality of our findings regarding AWAR from the empirical application. We then derive the CAWAR measure, which is similar to AWAR but adjusts for the cell size of a cohort, before testing it using Monte Carlo simulations. Section 8 explores alternative estimation techniques to pseudo-panels that can utilise cohorts as well as alternative techniques for estimating pseudo-panels that have certain useful properties. Section 9 concludes.

2 The Importance of Cohort Construction

This section details the reasons why cohort construction is such a crucial process to pseudo-panel estimation and why neglecting this issue may be problematic. In most applications of pseudo-panel estimation concerns regarding the way cohorts have been constructed essentially boil down to the singular issue of how large cell sizes (the number of individuals or households in each cohort group) are. This stems from Deaton's (1985) original paper where he identifies that unless cell sizes are large enough cohort dummies will not be sufficient to account for fixed effects present in a model. The reasoning is as follows.

Firstly, starting with the standard form of a model containing fixed effects:

$$y_{it} = \boldsymbol{x}'_{it}\beta + \theta_i + \varepsilon_{\text{it}} \qquad i = 1,2,...,\text{N} \quad t = 1,2,...,\text{T} \qquad (1)$$

Let y_{it} be the dependent variable for individual *i* at time *t*, $\boldsymbol{x}'_{it}$ is a vector of explanatory variables, and θ_i is the fixed effect. If we aggregate all *i* individuals belonging to cohort *c* at each time period *t* and then calculate the sample means of the cohorts, we can estimate the relationship:

$$\bar{y}_{ct} = \bar{\boldsymbol{x}}'_{ct}\bar{\beta} + \bar{\theta}_{ct} + \bar{\varepsilon}_{\text{ct}} \qquad c = 1,2,...,\text{C} \quad t = 1,2,...,\text{T} \qquad (2)$$

As there are additive individual fixed effects, there will be corresponding additive cohort fixed effects at the cohort level, shown by $\bar{\theta}_{ct}$. However, Deaton notes that these cohort fixed effects, as they are the average fixed effects of all individuals in each cohort, may not be fixed over time because in each time period there are different individuals in each cohort. Furthermore, as $\bar{\theta}_{ct}$ is unobserved and will in general be correlated with $\bar{\boldsymbol{x}}_{ct}$, cohort dummies cannot be used to account for the fixed effects. The only way cohort dummies could be used is if cell sizes are large enough that $\bar{\theta}_{ct}$ is a very good approximation of θ_c, the true cohort population fixed effect, and hence is essentially fixed over time. In this case one can estimate (2) using OLS with cohort dummies, which in the literature is known as the efficient Wald estimator (henceforth EWALD) following Angrist (1991). Deaton also recommends weighting the cohort level observations by the square root of the cell size (n_{ct}). The EWALD estimator is shown below where $\bar{\boldsymbol{x}}_{ct}$ includes the cohort dummies.

$$\boldsymbol{\beta}^{EWALD} = \left[\sum_{c=1}^{C}\sum_{t=1}^{T}(\sqrt{n_{ct}}\,\bar{\boldsymbol{x}}'_{ct}\bar{\boldsymbol{x}}_{ct})\right]^{-1}\left[\sum_{c=1}^{C}\sum_{t=1}^{T}(\sqrt{n_{ct}}\,\bar{\boldsymbol{x}}'_{ct}y_{ct})\right] \qquad (3)$$

If cell sizes are not sufficiently large so that $\bar{\theta}_{ct}$ can be considered as good approximations of θ_c and hence constant over time, Deaton proposes an alternative errors-in-variables estimator (henceforth EVE). For this an underlying unobserved cohort population version of equation (2) is proposed, for which the observed cohort sample means are conceived as error-ridden estimates of the true population means. As the variances and covariances of these sample means can be easily calculated from the survey data, EVE can be used to incorporate a measurement error correction using the results from Fuller (1975, 1981).

Most applied work has focused on constructing cohorts so that cell sizes are large enough that the simple EWALD estimator can be used rather than EVE, which has since been shown to be biased when T is small as it over-corrects for measurement error [see Verbeek and Nijman (1993)and Devereux (2007a)]. Naturally, the question then arises of exactly how large cell sizes need to be so that the simpler and more flexible EWALD estimator can be used. Most studies aim for cell sizes of at least 100 following Verbeek and Nijman (1992) where cell sizes of about 100-200 are found to be sufficient for EWALD estimates to be consistent and may be even smaller if the individuals grouped into cohorts have similar characteristics. Thus there is a general perception that as long as cell sizes are over 100, then using EWALD can deliver consistent estimates of linear fixed effects models when RCS data is available.

A slightly more nuanced view may also consider the bias versus efficiency trade-off that exists when grouping individuals into cohorts. The trade-off results from the fact that as the total number of individuals in the data (N) is fixed, cell sizes (n_{ct}) can only be increased by reducing the number of cohorts (C). Thus increasing the cell sizes may reduce the bias due to measurement error highlighted above, but at the cost of fewer observations at the cohort level thus reducing the efficiency of the estimates.

As a consequence one may be under the impression that cohort construction is a rather simple process where the main aim is to obtain cell sizes that are as large as possible while ensuring there are sufficient cohort observations for efficiency.

However a closer inspection particularly of the theoretical research implies otherwise. We discuss three important issues which complicate the cohort construction process. Firstly, the recommendation of having cell sizes of 100-200 requires some important caveats and should not be taken as a general standalone rule. Rather Verbeek and Nijman show that in order to mitigate the bias due to measurement error cell sizes need to be considered alongside variation present at the cohort level; in particular the true level of variance at the cohort level (w_1) and the measurement error variance of the observed cohort means (w_2). Where:

$$w_1 = \lim_{C\to\infty} \frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} (x_{ct}^* - \bar{x}_c^*)^2 \qquad (4)$$

$$w_2 = \operatorname{plim}_{C\to\infty} \frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} (\bar{x}_{ct} - x_{ct}^*)^2 = n_c^{-1} \sigma_v^2 \qquad (5)$$

x_{ct}^* are the unobserved cohort population means and $\bar{x}_c^* = \frac{1}{T} \sum_{t=1}^{T} x_{ct}^*$.

$\bar{x}_{ct}$ are the observed cohort sample means.

n_c is the cell size of cohort c.

σ_v^2 is the variance of individual x_{it} observations in cohort c.

The bias essentially depends on the relative size of w_1 and w_2 with the bias decreasing as w_1 increases compared to w_2. As cell sizes are just one component of w_2 thus consistency of EWALD will also depend on the size of w_1 and σ_v^2, both of which are affected by the way cohorts are constructed. Consequently, when one is constructing cohorts it is not sufficient to consider the size of the cells alone but also the degrees of variation being created at the cohort level.

The findings of Devereux (2007b) are even more humbling, suggesting that if there is insufficient variation amongst the cohort observations the cell sizes required may be in the thousands not hundreds, which is a condition very few datasets are likely to meet looking at existing studies shown in Tables 3.1a-c. Devereux uses US Current Population Survey (CPS) data to compare estimates when cell sizes of groups are

extremely large to estimates from randomly drawn subsamples from these larger groups. They estimated a model of female labour supply where eight cohorts were constructed (based on four 10-year birth cohorts and two education categories) tracked over 15 years, resulting in a total of 110 observations in the pseudo-panel with each cell having an average size of around 4500. Then they compared this estimate to those where 100, 200, 500, 1000 and 2000 observations were randomly drawn from the larger group, in each case carrying out 1000 replications. This was augmented by Monte Carlo simulation based on the above application where artificial data was created to mimic the CPS data. By doing so the author was able to set the values for the coefficients in the artificial data and compare them to the coefficient estimates. In this way they could accurately gauge the bias of the estimates.

The results from the actual data showed substantial bias in the EWALD estimates even with cell sizes of 2000. Furthermore, results from the Monte Carlo simulation indicated that even with 10,000 observations the bias was still of the order of 10%. However this is not a general finding as a similar model of male labour supply estimated in an earlier version of the paper found the bias of EWALD to be quite small once cell sizes reached 500. The reason for this discrepancy is because the variance over time for each cohort is too small compared to the variance across all cohorts within the same cross-section. It shows that the main problem is not the bias due to measurement error of group means, but the presence of cohort and year fixed effects that can cause serious small sample biases requiring thousands of observations to eliminate. In fact, the earlier working paper version contained additional Monte Carlo simulations which showed that omitting year dummies causes a substantial reduction in the median bias of EWALD estimates when cell sizes are 100; with the bias falling from 50% to about 10%. The larger bias is due to the year effects factoring out much of the across time variation in the data.

Thus both papers mentioned above highlight how having sufficient variation at the cohort level can reduce the cell size required for consistent EWALD estimates. Where they differ is that Verbeek and Nijman focuses more on the bias directly caused by measurement error and to eliminate this at most one or two hundred

observations are required for their dataset. Devereux uses a much broader framework and finds cohort and year fixed effects combined with measurement error causes small sample biases which are much more difficult to get rid of. For this reason it is difficult to arrive at a general set of rules which states the cell sizes required to get consistent EWALD estimates. The requirements differ for different datasets depending on the characteristics and level of variation generated amongst the cohorts being used. Thus one cannot just assume that because cohorts have cell sizes above 100 or 200 pseudo-panel estimates will be consistent.

Secondly, focusing on measurement error alone is not sufficient to produce consistent pseudo-panel estimates. For example, we have seen how having too little inter-cohort variation may exacerbate the bias caused by measurement error but if there is no genuine cohort level variation then the estimates may not have a well-defined probability limit at all. To demonstrate this (based upon Verbeek and Nijman, 1992 and Verbeek, 2008), consider the within estimator version of EWALD:

$$\boldsymbol{\beta}^{EWALD} = \left[\sum_{c=1}^{C}\sum_{t=1}^{T}(\bar{x}_{ct}-\bar{x}_{c})'(\bar{x}_{ct}-\bar{x}_{c})\right]^{-1}\left[\sum_{c=1}^{C}\sum_{t=1}^{T}(\bar{x}_{ct}-\bar{x}_{c})'(\bar{y}_{ct}-\bar{y}_{c})\right] \qquad (6)$$

The conditions for EWALD to be a consistent estimator are:

$$\operatorname{plim}_{n_c\to\infty}\frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}(\bar{x}_{ct}-\bar{x}_{c})'(\bar{x}_{ct}-\bar{x}_{c}) \qquad (7)$$

is finite and invertible and that

$$\operatorname{plim}_{n_c\to\infty}\frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}(\bar{x}_{ct}-\bar{x}_{c})\bar{\theta}_{ct}=0 \qquad (8)$$

To clarify matters, consider an example where cohorts are constructed based upon a variable that is independent of the variables in the model. Thus the true population cohort means for all cohorts would be the same for the variables as cohort assignment is essentially random. So x_{ct}^{*} (the true cohort population means of the explanatory variables) would be the same for all cohorts. If there was no variation

over time either, then $x^*_{ct} = \bar{x}^*_c$ for all *C* and all *T*. Thus the only variation left in the data would be down to measurement error, so $\bar{x}_{ct} - \bar{x}_c \neq 0$ is only due to measurement error. For example, if there are 10,000 individuals in the dataset who are randomly grouped into 100 cohorts, then one would expect the cohort means to be the same and only differ due to measurement error caused by insufficient cell sizes for the law of large numbers to hold. Furthermore, in this case increasing cell sizes is of little benefit since there is no genuine variation at the cohort level. Thus neither EWALD nor even EVE has a well-defined probability limit.

Hence for the conditions of finiteness and invertibility to be satisfied, there must be sufficient inter-cohort heterogeneity where the true cohort means vary across cohorts and over time. This often means that variables used to construct cohorts must in some way be related to the other variables in the model. For example, Moffitt (1993) proposes a very different approach to the estimation of RCS when fixed effects are present based on the idea that grouping methods are in fact instrumental variables methods. As such the cohort groupings must satisfy the exogeneity and relevance criteria of instruments. Thus for Deaton's within estimator of the pseudo-panel to be consistent, the cohort dummies interacted with time dummies are assumed to be sufficiently correlated with the explanatory variables and uncorrelated with the error term. This shows how important it is to consider the way cohorts are constructed and that one should be just as careful in constructing cohorts as in choosing instruments. Furthermore, one is limited by which cohorts can be used as they can only be defined by variables which are time invariant, exogenous and observed for all individuals in the sample, which imposes a severe restriction (Verbeek, 2008).

The third and final concern, related to whether there is sufficient genuine variation at the cohort level for the model to be identified, is how much variation is lost in the process of aggregation into cohorts. Many of the early studies which used pseudo-panel estimation were for the analysis of life cycle models of consumption and labour supply, see Table 3.1b for a summary. For such studies which constructed cohorts almost exclusively based on age groupings, analysis and inference at the cohort level was justified as behaviour over the life cycle was the main interest.

However, if one is interested more in analysing individual level behaviour rather than cohort level behaviour then one needs to question how applicable pseudo-panel estimation may be. So if one is using pseudo-panel estimation to complement cross-sectional estimates conducted at the individual or household level as a way to mimic panel estimates at that level, then one needs to be sure that $\overline{\beta}$ estimated in equation (2) is same as β from equation (1).

One reason as to why $\overline{\beta}$ and β may be different could result from the inherent bias caused by the act of grouping individual agents into a more aggregate level. Such aggregation may lead to the loss of important sources of variation and heterogeneity present at the individual level. Or one may also be concerned with problems similar to the fallacy of composition where behaviour at the aggregate level may be different and even contradictory to that at the individual level. Thus cohorts may need to be constructed at a sufficiently disaggregate level in order to allay such concerns.

The aim of this study is to look into these neglected concerns regarding cohort construction directly. The focus is on concerns regarding aggregation and the creation of genuine variation at the cohort level rather than focusing on cell sizes and measurement error alone, which have been addressed elsewhere in the literature, in order to show that the former is just as important as the latter and cannot be solved simply by having larger cell sizes or incorporating a measurement error correction into the estimation. We will be investigating directly how well different variables that can be used to construct cohorts perform at creating the requisite level of variation and hence are comparable to individual level estimates if genuine panel data was available. The ultimate goal is to produce some guidelines and rules in order to assess which cohort construction variables are most suitable and assess if the cohorts produced have the necessary level of variation for consistent estimation.

Furthermore, we will be focusing particularly on working with household or individual level data from developing countries. This is because within the field of development economics pseudo-panel estimation has become increasingly popular, see Table 3.1a for a sample of existing studies, and this trend is only likely to

continue with the expansion of the LSMS. Thus pseudo-panel estimation is increasingly being used to analyse household and individual level behaviour, making concerns regarding the aggregation into cohorts and their comparability with individual level estimates more pressing.

3 Data and Methodology

3.1 Data

The dataset we use in order to investigate how cohort construction affects pseudo-panel estimation comes from the Uganda National Panel Surveys (UNPS) which contains four rounds taken in 2005/06, 2009/10, 2010/2011 and 2011/2012. The original 2005/06 data is taken from the Uganda National Household Survey which contained a nationally representative sample of about 7,400 households from 783 enumeration areas. The panel over the next 3 waves is constructed by reinterviewing 3,123 households from 322 of the original 783 enumeration areas located all over the country. However as not all households were able to be reinterviewed there is some attrition over the waves; with each successive wave containing 3123, 2607, 2564 and 2356 households respectively. The dataset contains detailed information on household consumption, income, wealth, labour market activities as well as information on the characteristics of individuals in the household such as their age and education level.

3.2 Methodology

The dataset used is ideal for our purposes as it contains detailed individual and household level data which can be used to estimate models that are then compared to cohort level estimates. The dataset was also constructed under the LSMS and shares similar attributes to data available for other developing countries. This means the variables and the methods we use to construct cohorts can be replicated by others using similar data. The panel nature of the dataset is crucial for a number of reasons.

Firstly, it will give us the "true" genuine panel estimates from a linear fixed effects model to which we can compare the pseudo-panel estimates. Thus we can compare how alternative ways of constructing cohorts lead to estimates closest to the genuine individual level panel results. If we use RCS data, we could only compare pseudo-panel estimates to either cross-sectional estimates or other pseudo-panel estimates. The cross-sectional estimates will be biased if fixed effects are present thus will not provide a good comparison to assess the reliability of the pseudo-panel estimates. Comparing different pseudo-panel estimates to each other would also not prove useful as we do not know whether and what extent each estimate suffers from bias arising from the aggregation process.

Secondly, one concern with using panel data is that generally they have much smaller samples due to the cost and difficulty of tracking the same large number of individuals over time. Thus say we have a reasonably large panel of 4000 individuals traced over 4 years. We can use fixed effects estimation to get the "true" estimates to compare to pseudo-panel estimates. To estimate pseudo-panel models one would then need to get rid of the panel nature of the data so that it reflects the RCS that pseudo-panel estimation is used for. Doing so would leave us with 1000 unique individuals in each time period to work with. Given cell sizes may need to be at least 50-100 in order to account for the measurement error issue, thus the number of cohorts we could use would be at most 10-20, which is not a sufficient number for us to fully investigate the issues arising from aggregation. In order to do that, we would ideally want to be able to construct hundreds of cohorts so we can compare estimates at vastly different levels of aggregation. This would require having data with at least 5000-10000 unique individuals in each time period. Generally it would be very difficult to find panel data which fit these criteria.

However for our purposes there is a simple solution. As we are focusing specifically on bias and identification issues arising just from aggregation into cohorts and not on measurement error for cell sizes, by using the panel nature of the data we can mitigate the measurement error problem even if cell sizes are very small. Recall that the measurement error bias arises from the fact that $\bar{\theta}_{ct}$ is not constant over time as the individuals in a cohort change over time when using RCS data, hence cohort

dummies cannot account for fixed effects. However, when panel data is used the individuals being grouped into cohorts are fixed over time thus $\bar{\theta}_{ct}$ will also be fixed over time irrespective of cell sizes. At the extreme, even if cell sizes are as small as two, the cohort fixed effects remain constant over time as the same two individuals make up the cohort for all time periods. This means in our hypothetical panel dataset of 4000 individuals over 4 waves, we could use all 4000 individuals in each time period and we could create up to 2000 cohorts (each with a cell size of 2) without having to worry about measurement error. This means we can compare pseudo-panel estimates based on 10-20 cohorts to others based on 1000-2000 cohorts as well as to the "true" panel estimates, allowing us much freedom to explore the aggregation issue.

In this way the UNPS data can be used to explore aggregation issues arising from the process of cohort construction. Furthermore cohorts can be created in a fashion replicable by others within the field of empirical development economics with access to similar data, particularly produced under the LSMS. At this stage it would be useful to see how cohorts may and have been constructed. Tables 3.1a, 3.1b and 3.1c detail cohort construction criteria used in existing empirical research where we group papers respectively into those for developing economies using household or individual level data, those estimating life cycle models, and other papers which use household or individual level data from developed economies, often estimating the demand for a certain good. The papers estimating life cycle models generally have a small number of cohorts and large cell sizes as they mainly construct cohorts purely based on age, mostly 5 year age cohorts. As we mentioned before, this is due to the very nature of the models where age cohorts themselves are the main focus of the analysis.

Looking at the other two categories of studies, in Tables 3.1a and 3.1c, we see that the number of cohorts is generally larger but cell sizes are often smaller, especially when other variables are combined with age to produce cohorts, particularly in the development economics literature which also has a number of studies where age is not used at all to produce cohorts. Across all studies mostly 5 year age cohorts are used, but there is quite a broad spectrum of age bands used from 1 and 2 years to 10

years. Geographical characteristics such as the region, state or district individual/households are from is the most often used variable after age to construct cohorts. Education is also quite popular and it should be noted that the studies which use education do so by generally categorising individuals into about two or three groups, such as whether the individual has compulsory or post-compulsory education in a developed country setting or if they have no education, primary or post-primary education in a developing country setting. When household level data is used education level is based on the education of the household head. Other variables used to construct cohorts include the gender of the individual or household head, their location (whether they come from a rural or urban area) and their ethnicity. Although more common in studies outside the development literature, some papers do use variables related to the socioeconomic conditions to construct cohorts. This includes socioeconomic status directly, whether individuals are homeowners or renters, the type of employment they are engaged in or some measure of their income or wealth.

We will be focusing on using construction variables that are most likely to be available and relevant for researchers using LSMS data to make our findings as general as possible. Recall some important conditions for construction variables are that they are time invariant, exogenous and observed for all individuals in the sample. Thus we exclude the use of the socioeconomic variables as these are likely to be endogenous and have individuals or households moving in and out of these categories over time in developing economies, especially when the reason behind using RCS data is to cover a large time span. We also exclude ethnicity from our list of possible variables as this is something likely to be relevant to only a few countries and in fact is not available in the UNPS data. Thus we will be using five potential variables to construct cohorts: the age of the household head, their gender, education level, the region of the household and whether it is in a rural or urban location.

Table 3.1a: Pseudo-panel Estimation in Development Economics

Article	**Cohorts**	**Cell size**	**Cohort construction**
Bedi et al. (2004)	38	350	Districts in Kenya
Christiaensen and Subbarao (2005)	799	10	Communities in Kenya
Nicita (2009)	63		States in Mexico, location (urban or rural)
Antman and McKenzie (2007a)	15	100+	5 year age cohorts, education
Warunsiri and McNown (2010)	22 11	200+ 300+	1 year age cohorts 2 year age cohorts
Cuesta et al. (2011)	224	130	7 year age cohorts, gender, country in Latin America
Sprietsma (2012)	108	130	States in Brazil, gender, ethnicity
Échevin (2013)	166	115.5	5 year age cohorts, education, region
Fulford (2014)	200+		5 year age cohorts, states in India, gender
Imai et al. (2014)	140	73.6	5 year age cohorts, states in India
Himaz and Aturupane (2016)	21 11	318 608	1 year age cohorts 2 year age cohorts
Gómez Soler (2016)	6000+	70-80	Schools in Colombia

Table 3.1b: Pseudo-panel Estimation of Life Cycles models of Consumption and Labour Supply

Article	Cohorts	Cell size	Cohort construction
Browning, Deaton and Irish (1985)	16	192	5 year age cohorts, manual vs non-manual worker,
Banks, Blundell, and Preston (1994)	11	354	5 year age cohorts
Blundell, Browning, and Meghir (1994)	9	520	5 year age cohorts
Deaton and Paxson (1994)	56 14 11	300-400 200-400 150-200	1 year age cohorts for Taiwanese data 5 year cohorts for US data 5 year cohorts for British data
Alessie, Devereux, and Weber (1997)	5	250+	10 year age cohorts
Blundell, Duncan, and Meghir (1998)	8	142	10 year age cohorts, education
Fernandez-Villaverde and Krueger (2007)	10	350	5 year age cohorts
Attanasio et al. (2009)	15	500+	5 year age cohorts

Table 3.1c: Pseudo-panel Estimation in General Models of Individual Behaviour

Article	Cohorts	Cell size	Cohort construction
Gassner (1998)	27	226	2 year age cohorts
Dargay and Vythoulkas (1999)	16	513	5 year age cohorts
Propper et al. (2001)	70	80	5 year age cohorts, region
Dargay (2002)	41	190	5 year age cohorts, location
Gardes et al. (2005)			10 year age cohorts, education
Campbell and Cocco (2007)	7 9 12	200+ 150+ 100+	5 year age cohorts 10 year age cohorts, region 5 year age cohorts, if homeowner or renter
Bernard et al. (2011)	25	131	Region, Size of house
Jiang and Dunn (2013)	15		5 year age cohorts
Meng et al. (2014)	72	140	5 year age cohorts, gender, socioeconomic status, region

3.3 Model and Results from Panel Estimation

The next section details the specifics of how these construction variables will be used to create cohorts from which pseudo-panel estimates will be derived. For now, we will outline the model to be used in order to compare and analyse the consistency of the pseudo-panel estimates derived from the different construction methods. To do this we set up a model of household welfare where welfare is measured by the natural logarithm of monthly household consumption per adult equivalent (labelled *lcons*) member following Deaton and Zaidi (2002). This model is chosen because it is a useful baseline model for many empirical studies using such household data and fixed effects are likely to be present due to unobserved factors like preferences, intra-household bargaining process and innate characteristics of household members. The model to be estimated is shown below and is a general model of household consumption that can be commonly found in the related literature, see Glewwe (1991) and Appleton (1996) for examples:

$$\begin{aligned} lcons_{it} &= \alpha + \beta_1 \boldsymbol{HHsize} + \beta_2 \boldsymbol{ltotasset} + \beta_3 \boldsymbol{secondary} + \beta_4 \boldsymbol{lnremit} \\ &+ \beta_5 remittance + \beta_6 age + \beta_7 age^2 + \beta_8 education + \beta_9 urban + \beta_{10} female \\ &+ \beta_{11} region + \beta_{12} \text{sec}tor + f_i + f_t + e_{\text{it}} \end{aligned} \qquad (9)$$

where *HHsize* is the size of the household, *ltotasset* is the log of total assets owned by the household not including agricultural and enterprise assets, *lnremit* is the log of remittances received by the household in the past year (we also include a dummy variable indicating whether the household received any remittance), *secondary* is a dummy variable indicating whether the household head had a secondary source of income in the past year, *age* and age^2 are the age of the household head at the time of the survey and its square, *education* indicates the education level of the household head which can be either no education or primary education or post-primary education, *urban* is a dummy that equals one if the household is located in an urban area as opposed to rural, *female* is a dummy indicating whether the household head is female. *Region* indicates whether the household is in one of four regions within Uganda, *sector* indicates the sector of the household head's main economic activity (self-employed agriculture, self-employed non-agriculture, government sector,

private sector agriculture or non-agriculture or unemployed), f_i indicates the individual effects potentially correlated with some of the regressors and f_t represents time effects.

Table 3.2 shows the results of estimating the model in equation (9) using pooled OLS, random effects, and fixed effects estimation. The Breusch and Pagan Lagrangian multiplier test and the robust version of the Hausman test imply that the correct estimation to use is fixed effects. We choose to focus on the four regressors highlighted in bold in equation (9). These variables are selected as they are of differing forms and react differently to the presence of unobserved heterogeneity. The forms differ in that household size is a discrete count variable, log of total assets is a continuous variable, secondary occupation is binary and the log of remittances received is a censored regressor with some continuous values and others that are zero. To this we include the remittance dummy more as a control to allow for a separation between the effect of receiving remittance at all and the effect of an increase in the actual amount of remittance. Summary statistics for the variables of interest are presented in Table 3.A1 of the Appendix for the first wave, last wave and full dataset.

In terms of reaction to presence of unobserved heterogeneity, we see that household size is the least affected: the OLS estimates, once controlling for additional explanatory variables, are very similar to the random and fixed effects estimates in terms of size and significance. The patterns for the log of total assets, log of remittances and the remittance dummy are all very similar in that coefficient estimates decrease in magnitude as more controls are added, as random effects are accounted for (unobserved heterogeneity that is assumed to uncorrelated with the regressors) and as fixed effects are used (relaxing the assumption of no correlation between regressors and the unobserved heterogeneity). However the sign and significance level of all three variables stay largely the same across all specifications. The log of total assets sees the smallest relative change in magnitude as the specifications change although the absolute value of the changes is similar to those witnessed by the log remittances. Although it may seem that the remittance dummy has great fluctuations in the magnitude of its estimates, one should note that these are

of roughly the same multitude as the fluctuations in the log remittance estimates and that the former estimates are always roughly 10-12 times the size of the latter.

Table 3.2: Estimation of full model of household consumption

	A OLS	B OLS	C RE	D FE	E FE	F FE
HHsize	-0.116***	-0.091***	-0.090***	-0.091***	-0.088***	-0.090***
	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
ltot_asset	0.207***	0.149***	0.125***	0.075***	0.077***	0.100***
	(0.008)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)
secondary	-0.082***	-0.031***	0.000	0.050***	0.058***	0.048***
	(0.014)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)
lnremit	0.136***	0.074***	0.057***	0.027***	0.024***	0.026***
	(0.008)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
remittance	-1.581***	-0.810***	-0.622***	-0.279***	-0.256***	-0.283***
	(0.090)	(0.077)	(0.071)	(0.080)	(0.080)	(0.080)
Other controls		Yes	Yes	Yes		
Observations	9,966	9,963	9,963	9,963	9,966	9,510
R-squared	0.429	0.608	0.414	0.435	0.430	0.432
Number of Households			3,413	3,413	3,413	3,342

Note: We report the within-R^2 for RE and FE regressions. All regressions contain year dummies to allow for year fixed effects. Other controls include all explanatory variables in equation (9). The last column trims at the top and bottom 1 percentiles to limit the impact of outliers. Columns A and E have 3 more observations as the inclusion of some additional control variables leads to omission of some households. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The variable most affected by the allowance for fixed effects is the secondary job dummy. Under OLS secondary occupation has a negative and significant effect both when using the full model with all controls or the more parsimonious model with just the main variables of interest. Once allowing for random effects the effect of secondary occupation becomes insignificant, however with fixed effects it becomes highly significant again but now positive. This positive effect is more intuitive theoretically as one would assume that ceteris paribus the household head having a secondary occupation would increase household incomes and thus allow for greater consumption.

Thus one can see a spectrum of how allowing for fixed effects alters the coefficient estimates of the regressors. At one end we have household size which is very little

changed, followed by log total assets and the remittances variables where sign and significance stay the same but the magnitudes decrease. Secondary occupation is the most affected, changing from negative and significant under OLS to positive and significant under FE. The form of the regressor does seem to be important, with quite stable estimates for count or continuous variables but unstable estimates for coefficients on binary variables. This has implications for how cohort construction and pseudo-panel estimation are differentially affected by the type of regressor. One could reasonably assume for example that pseudo-panel estimation would perform worse for binary variables compared to continuous variables because there is less variation in the former. Additionally one could assume that the more important individual fixed effects are for a regressor in the panel results, the more important the identification of cohort effects in the pseudo-panel results will be and thus the more important is the bias from aggregation as cohort effects are used to replace the individual effects.

In order to assess the different cohort construction methods to see which performs the best we will be comparing the pseudo-panel estimates from each construction method to the coefficient estimates in Column F of Table 3.2. We choose this specification as the baseline or "true" coefficient estimates rather than the other two fixed effects specification in Columns D and E for two main reasons. Firstly, unlike the specification in column D, it uses a more parsimonious model where only the main explanatory variables are included with year effects and no other additional controls. This is beneficial as some of those controls would naturally need to be excluded in pseudo-panel estimation as they will be used to construct the cohorts. Also, having a parsimonious model is useful at the cohort level where the number of cohort level observations will be far lower and we may be left with as few as 30 to 40 observations in total depending on how cohorts are constructed. In addition, as can be seen from comparing the results from Columns D and E, estimates are largely unaffected when these other control variables are removed, thus there is little loss from parsimony. Secondly, we use the parsimonious model with trimming of outliers at the 1st and 99th percentiles, where the impact of trimming can be assessed by comparing Columns E and F. We trim in order to reduce the effect of outliers which is especially important when comparing to pseudo-panel estimates where outliers

naturally have less impact due to the averaging of values across individuals in each cohort. It would also be useful to compare the pseudo-panel estimates to the results in Column B as when using RCS data the genuine alternative to pseudo-panel estimation is pooled OLS with controls for observable variables. As such, seeing if the pseudo-panel estimates are closer to the fixed effects estimates than pooled OLS estimates gives a useful way to assess whether they offer an improvement to OLS.

4 Pseudo-Panel Results

Looking at Table 3.1a, which contains studies using similar data, one can see there is no standard approach to constructing cohorts in the literature and thus little guidance on how it should be done. In this section we use the five cohort construction variables available, discussing how they can be used to create cohorts in practice, and applying them using an exhaustive approach which considers all the possible ways of aggregating into cohorts. This results in 65 different options available and we estimate pseudo-panel models with all of them. We find a great deal of variation amongst the estimates, where some are as good as the panel FE estimates while others are far worse than using OLS, highlighting the crucial importance of cohort construction to pseudo-panel estimation. We also show that many of the common methods of constructing cohorts in the literature are unable to deliver accurate estimates in our application.

When constructing cohorts age is quite ubiquitous in the applied literature, particularly when using household or individual level data. Cohorts are defined based on an age or year of birth band. Various bands have been used in the literature, the most common being 5 year bands however others have also used 1 year, 2 year and 10 year bands. The smaller bands are generally used more commonly when looking at individual level data, for example to calculate returns to education, however there is little guidance about what the optimal age band should be. We investigate this issue by constructing cohorts based on 2 year, 5 year, 10 year and 17 year age bands, giving us a good range of bands to assess. Two year age band is used

rather than one year as the survey was conducted over the period of two calendar years thus one year age bands are likely to be more erroneous. The 17 year age band essentially allows for the splitting of households into three categories; young, middle-aged and old. We use households whose head are aged 18-67 at the time of survey for each round. The distribution of households by household head age can be seen in Figure 3.A1 of the Appendix and summary statistics for age broken down by each survey round can be found in Table 3.A2 of the Appendix.

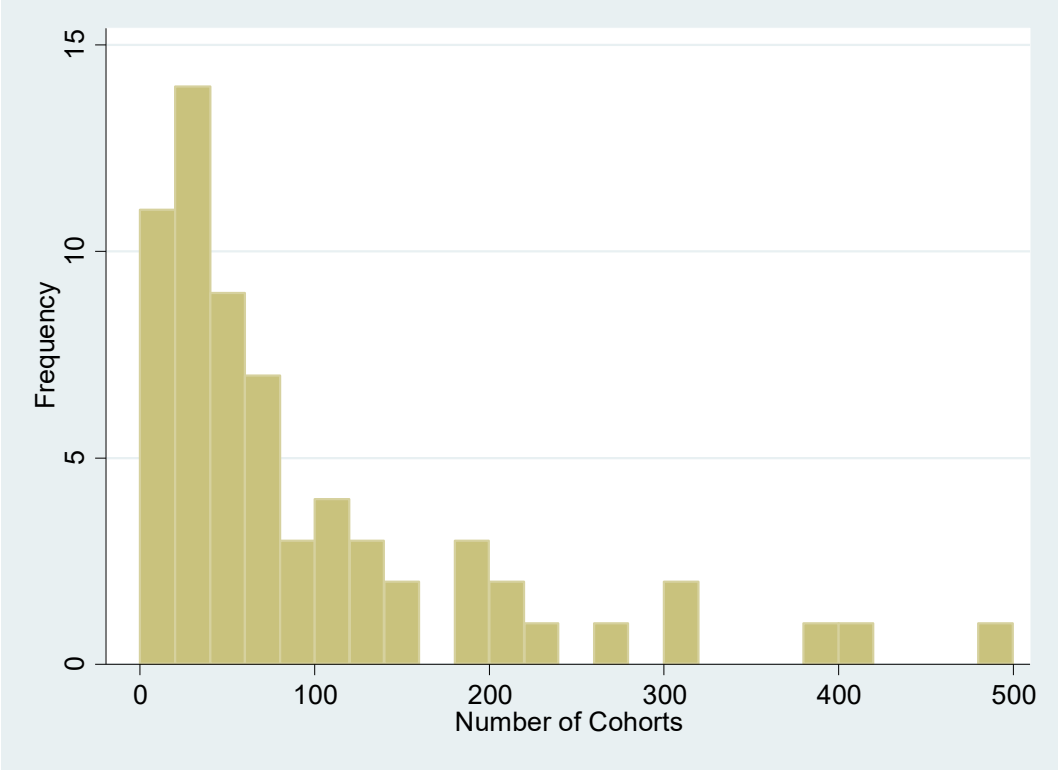
Table 3.3: Cohort Construction Variables

Variable	**Categories**	**Number of Observations**	**Percentage of Total**
Region	Central	3101	31.1
	Eastern	2307	23.2
	Northern	2442	24.5
	Western	2116	21.2
Gender	Male	7174	72.0
	Female	2792	28.0
Location	Rural	7540	75.7
	Urban	2426	24.3
Education	None	1516	15.2
	Primary	5347	53.7
	Post-primary	3103	31.1

The other four construction variables can be combined with age cohorts in 16 different ways (4 individually, 6 in pairs, 4 combinations of three variables, one which contains all four variables and another one where none are used). Table 3.3 shows the number of categories these four grouping variables are divided into as well as the number and proportion of households in each category. Combining these 16 different ways with the four different age bands described above gives us 64 potentially different grouping methods to choose from. To this we can add another 15 methods where age is not used but just different combinations of the other four construction variables. This gives a total of 79 different ways cohorts may be constructed. However not all of the 79 are viable because some will produce too few observations at the cohort level to run pseudo-panel regressions. For example, combining the 17-year age bands (3 categories) with gender (2 categories; male and female) gives just 6 cohort level observations for each time period and 24

observations in total. As we not only need to include time dummies and cohort dummies in the estimation, the number of observations is insufficient for efficient estimation. Thus we ignore all grouping sets that lead to fewer than 40 cohort level observations in total (i.e. at least 10 per wave), which still leaves us with a total of 65 potential cohort construction methods to choose from. The histogram in Figure 3.1 shows the dispersion of the number of cohorts per wave created across the 65 methods. There is a wide spectrum of how aggregated the cohort groupings are, with the number of cohorts for each wave being as small as 10 to as large as nearly 500. This will be crucial in identifying any biases or problems arising from the degree of aggregation as well as seeing how individual construction variables perform.

Figure 3.1: Distribution of the number of cohorts



About 70 per cent of households are observed in all four time periods. If we only use these households we would ensure that households within a cohort are fixed over time and the measurement error problem is restricted. However, this is likely to cause problems due to attrition both in the panel and pseudo-panel estimates.

Furthermore, it may be useful to keep a certain degree of measurement error because problems arising from aggregation and the lack of variation at the cohort-level are known to interact with and extenuate the measurement error bias. Thus one cannot grasp the true extent of these problems unless some measurement error remains present. Therefore we keep all households in the dataset even if they only appear in one survey round. As the vast majority of households appear in all four periods this means measurement error will not be so big an issue that it will overshadow the other sources of bias we are interested in, nonetheless it will still be present to a sufficient degree in order for it to interact with those other sources of bias, thereby ensuring a more realistic setup is being used.

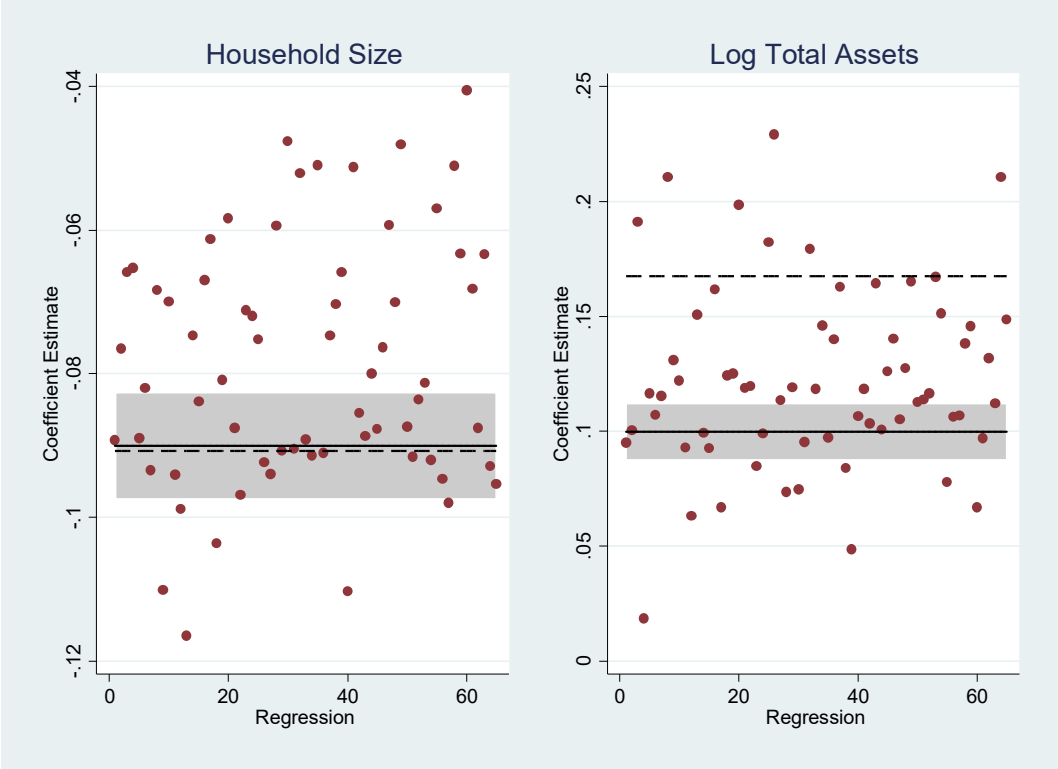
If these 65 different methods give similar estimates then one can be confident that cohort construction has little impact upon the reliability of pseudo-panel estimates and the bias vs efficiency trade-off (having larger cell sizes against more cohorts) is the main concern one needs to address when thinking of grouping individuals into cohorts. However, if there are large differences between the estimates then this is an indication that cohort construction needs to be of greater concern due to potential aggregation bias and insufficient cohort-level variation. We therefore use all 65 construction methods and estimate with Deaton's EWALD estimator. Cohort-level observations are weighted by the square-root of the cell size and we drop all cohorts that contain less than two households to ensure we retain the grouped element and pseudo-panel results are not dominated by cohorts which essentially identify a single household thus making it more akin to genuine panel estimation. Once cohorts have been constructed for each of the 65 methods we then estimate the model given in Equation (10), which is the cohort version of the model in Equation (9) used for panel estimation but excluding the additional controls.

$$lcons_{ct} = \alpha + \beta_1 \boldsymbol{HHsize} + \beta_2 \boldsymbol{ltotasset} + \beta_3 \boldsymbol{secondary} + \beta_4 \boldsymbol{lnremit} + \beta_5 remittance + f_c + f_t + e_{ct}$$

Figures 3.2a and 3.2b graph the coefficient estimates for each of the 65 construction methods. We exclude the results for the remittance dummy as firstly it is mainly added as a control for the variable we are really interested in, log remittances, and

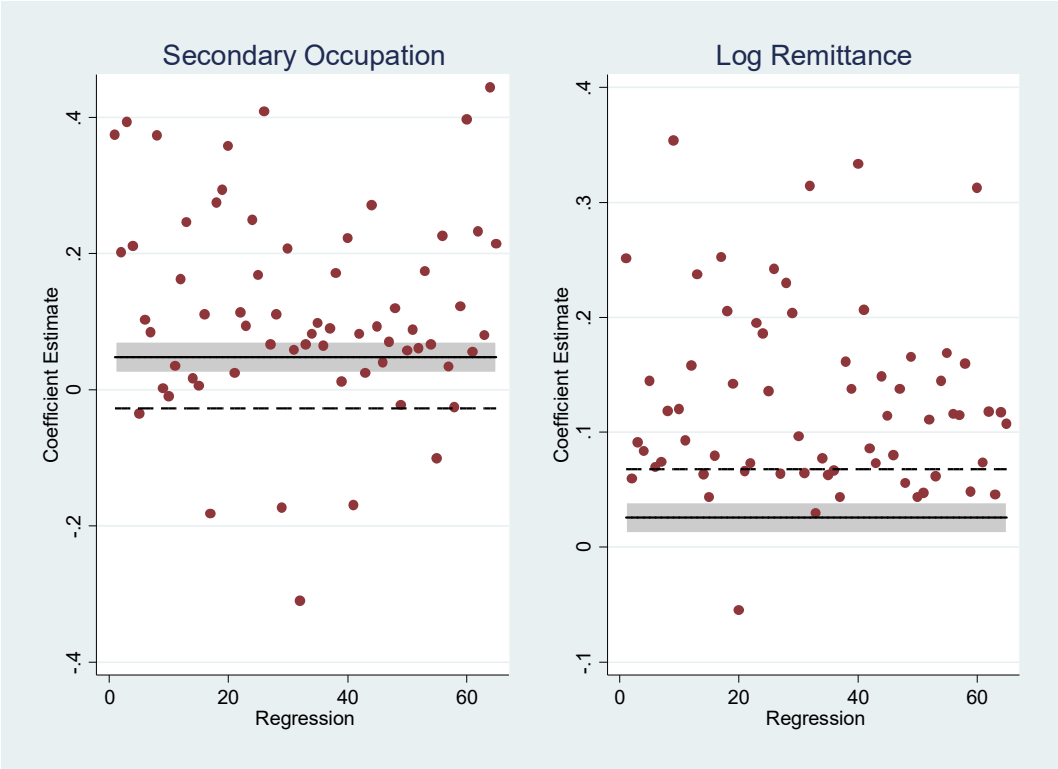
secondly because it largely follows the same pattern as log remittances (the coefficient estimates for both variables are closely related). The solid lines are the baseline or "true" coefficient values, from the fixed effects estimates excluding additional control variables and trimming the data at the 1st and 99th percentiles (Column E of Table 3.2), against which we compare the pseudo-panel estimates. The shaded area around the solid lines is the 95% confidence intervals of the "true" values. The dashed lines are the OLS estimates, including the other control variables with trimming (Column A of Table 3.4a). We include these values as well because when one has RCS data the main alternative to pseudo-panel estimation is pooled OLS, thus by comparing the two one can see if pseudo-panel estimation can offer an improvement.

Figure 3.2a: Pseudo-panel estimates from all 65 Cohort Specifications



Note: The scatter plots are the 65 pseudo-panel estimates obtained by the different cohort construction methods. The solid line shows the "true" fixed effects estimates with the grey area being its 95% confidence interval and the dashed line represents the OLS estimates. Both estimates are derived from the household level panel data with the data trimmed at the 1st and 99th percentiles.

Figure 3.2b: Pseudo-panel estimates from all 65 Cohort Specifications



See footnote for Figure 3.2a.

For all four variables considered we see a large amount of variation in the pseudo-panel estimates, showing the importance of cohort construction as the only difference between the estimates are how cohort aggregation is conducted. The pseudo-panel estimates can range from being very close to and as good as genuine panel estimation to being far worse than even OLS. This latter issue, of being worse than OLS, highlights the point that the reason many pseudo-panel estimates are performing poorly is not because the cohort effects are not accounting for the unobserved heterogeneity but instead are coming from biases due to the aggregation process itself and potentially due to the loss of variation at the cohort level. This is best typified by the household size estimates shown in the left panel of Figure 3.2a. Here we see that the OLS and FE estimates are almost identical, meaning accounting for individual heterogeneity is not necessary. Nevertheless most of the pseudo-panel estimates perform rather poorly, being quite far away from the "true" FE estimate and its 95% confidence interval. A similar problem can be seen for log remittances estimates in the right panel of Figure 3.2b. Here one can see that most pseudo-panel

estimates perform worse than OLS, containing a larger upward bias than what is due to the presence of fixed effects alone.

For log total assets we see that accounting for unobserved heterogeneity is necessary as the FE estimate is somewhat smaller in magnitude than the OLS estimate. The pseudo-panel estimates in general outperform the OLS estimates, due to the inclusion of cohort effects which picks up the heterogeneity, and many of the estimates are very close to the "true" FE values. However there still remains a significant portion that offer little improvement over OLS or are worse, showing that a reduction in bias from the inclusion of cohort effects can be offset by increased bias coming from aggregation. One can see a similar pattern for secondary occupation in Figure 3.2b, for which including fixed effects is most important as OLS indicates a negative and significant coefficient whereas FE results in a positive and significant coefficient. The pseudo-panel estimates pick up the effect of fixed effects with the vast majority of coefficient estimates being positive, however many are still highly biased as the magnitudes are often five to ten times larger than the "true" FE values. It also gives a good example of just how much impact cohort construction has as the coefficient estimates can vary from being smaller than -0.3 and as large as over +0.4 even though the FE and OLS estimates are +0.05 and -0.03 respectively.

Looking into some specific pseudo-panel estimates, we first consider some cohort construction methods that can be commonly found in the literature, four of which can be seen in Table 3.4a. Using just age cohorts alone is very popular in the applied literature, especially for studies estimating life cycle models where 5 year age bands are most often used. In the development economics literature, where the analysis is more at the individual level than at cohort level, if solely age cohorts are used then often shorter bands such as 1 or 2 years is more common (see Table 3.1a). For example, Warunsiri and McNown (2010) and Himaz and Aturupane (2016) both look at returns to education and use 1 and 2 year bands. In Column C we present the results from when 2 year age bands alone are used to construct cohorts. One sees that the results are not very promising with the OLS estimates often performing better. The household size coefficient is relatively small in magnitude, the log total assets

estimate is largely the same as the OLS estimate, although the secondary occupation coefficient is positive it is about twice the size of the FE value and also insignificant, and the remittance variables are larger in magnitude than even the OLS estimates as well as being insignificant.

Another common cohort aggregation method is to combine age cohorts with regions of the country, thus in Column D we combine the 2 year age bands with region. Again one sees little indication that pseudo-panel estimation is worth carrying out rather than just relying on OLS. The household size estimate is fairly accurate however the log total assets estimate is now even larger and more biased than OLS. Secondary occupation is positive and significant, like the "true" FE coefficient, but in terms of magnitude it is over three times larger than the "true" coefficient. Similarly, even though the remittance variables retain their significance the magnitudes of the coefficient far exceed not just the FE estimates but even the OLS.

In the last two columns of Table 3.4a we look into the common assertion that as long as cell sizes are 100-200 then pseudo-panel estimates will be quite reliable. In Column E the average cell size of the cohorts is 124 with the cohorts based on 5 year bands and the location of the household (urban or rural), both of which are common variables used in the literature. The same can be said of region and education which we use to construct cohorts in Column F, where the average cell size is 208, so relatively large for cell size as identified in the literature. We find both these construction methods provide very poor estimates for all the explanatory variables, far worse even than just using simple Pooled OLS. Thus we see that many of the common methods and heuristics of constructing cohorts from the literature have failed to not only produce pseudo-panel estimates which are similar to FE results but are actually even worse than OLS.

The estimates in Table 3.4a are just a few of the many pseudo-panel estimates which perform quite poorly from our selection of 65. However, there are number of estimates which perform quite well and give results very similar to the FE estimates, three of which we have included in Table 3.4b. We see that the pseudo-panel results are very similar in sign, significance and size to the FE estimates for all the

independent variables, meaning that not only has the individual heterogeneity been accounted for but also there is little variation lost and hence less bias arising from the aggregation process.

Table 3.4a: Pseudo-panel estimates based on common cohort specifications

	Pooled OLS	Fixed Effects	Pseudo-Panel 1	Pseudo-Panel 2	Pseudo-Panel 3	Pseudo-Panel 4
HHsize	-0.093***	-0.090***	-0.067***	-0.075***	-0.061***	-0.040
	(0.003)	(0.004)	(0.015)	(0.014)	(0.020)	(0.088)
ltot_asset	0.167**	0.100***	0.162***	0.182***	0.067	0.067
	(0.005)	(0.006)	(0.038)	(0.025)	(0.060)	(0.076)
secondary	-0.027***	0.048***	0.110	0.168*	-0.183	0.397**
	(0.011)	(0.012)	(0.162)	(0.091)	(0.145)	(0.171)
lnremit	0.068***	0.026***	0.080	0.136***	0.252***	0.313***
	(0.006)	(0.007)	(0.057)	(0.036)	(0.079)	(0.109)
remittance	-0.747***	-0.283***	-0.893	-1.504***	-3.546***	-3.412***
	(0.074)	(0.080)	(0.688)	(0.425)	(1.00)	(1.227)
Other Controls	Yes	No	No	No	No	No
Observations	9507	9,510	9,855	9,837	9,939	9,966
Cohort Constructors			2 year	2 year, Region	5 year, Location	Region, Education
Cohorts			25	96	20	12
Cell Size			100	26	124	208

Note: The fixed effects and OLS regressions trim the top and bottom one percentile. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.4b also shows that the variables which have the closest pseudo-panel estimates to the "true" FE values both in terms of size and significance are household size and log total assets. Secondary occupation and the remittance variables lose some significance in the pseudo-panel estimates and the size of the coefficients, particularly for the remittance variables, can be somewhat larger too. This pattern can more generally be seen by comparing the variables in Figure 3.2a to those in Figure 3.2b. One sees that household size and log total assets have more estimates close to the "true" FE estimate and its 95% confidence interval than secondary occupation and log remittance. Further evidence of this can be found in Table 3.5, where we give some summary statistics of the 65 estimates for each regressor. One sees that household size and log total assets have mean and median values which are

closer to the "true" FE estimates as well as a smaller mean absolute bias (thus are more accurate) and have smaller standard deviation and range than the other two regressors. Thus it appears pseudo-panel estimates of these two variables are on average less biased and more precise (have lower variance).

Table 3.4b: Accurate Pseudo-panel estimates

	Pooled OLS	Fixed Effects	Pseudo-Panel 4	Pseudo-Panel 5	Pseudo-Panel 6
HHsize	-0.093***	-0.090***	-0.087***	-0.089***	-0.090***
	(0.003)	(0.004)	(0.007)	(0.007)	(0.009)
ltot_asset	0.167***	0.100***	0.112***	0.118***	0.095***
	(0.005)	(0.006)	(0.013)	(0.012)	(0.015)
secondary	-0.027**	0.048***	0.057*	0.067**	0.058
	(0.011)	(0.012)	(0.034)	(0.031)	(0.040)
lnremit	0.068***	0.026***	0.044**	0.030*	0.065***
	(0.006)	(0.007)	(0.018)	(0.015)	(0.020)
remittance	-0.747***	-0.283***	-0.517**	-0.332*	-0.747***
	(0.074)	(0.080)	(0.215)	(0.181)	(0.241)
Other Controls	Yes	No	No	No	No
Observations	9966	9,510	9407	8583	9543
Cohort Constructors			2 year, Region, Location, Education	2 year, Region, Location, Education, Gender	5 year, Region, Location, Education, Gender
Cohorts			363	435	291
Cell Size			6.5	5	8

Note: The fixed effects and OLS regressions trim the top and bottom one percentile. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.5: Summary statistics of 65 Pseudo-panel estimates

	Household Size	**Log Total Assets**	**Secondary Occupation**	**Log Remittance**
"True" Value	-0.090	0.100	0.048	0.026
Mean	-0.079	0.122	0.115	0.127
Median	-0.081	0.117	0.090	0.114
Standard Deviation	0.017	0.040	0.149	0.081
Mean Absolute Bias	0.016	0.033	0.119	0.103
Range	0.076	0.211	0.754	0.409

This is an important point to note because looking at the two sets of variables the ones that perform better have a more continuous distribution in their form. Log total assets is a continuous variable and household size, although a discrete variable, still has a wide distribution varying from as small as 1 to as large as 20 (or 14 when trimmed at 99th percentile). Of these two explanatory variables, generally household size is better estimated (Table 3.5) and this is likely due to it being the least affected by individual heterogeneity. Of all four regressors the worst performer is secondary occupation, particularly in terms of the large standard deviation and range of estimates produced. Again this seems to correspond to the fact that it is the least continuous variable, being binary, as well as suffering greatly from the presence of heterogeneity.

A further point illustrated by these results is that just because one regressor is well estimated by pseudo-panel estimation does not imply the other explanatory variables in the model will be well estimated as well. For example one may anticipate that household size is uncorrelated with unobserved heterogeneity, hence the OLS estimate will be unbiased (i.e. they anticipate the similarity between the OLS and FE estimates). Thus one may think that a good way to assess how well cohort construction has been conducted is by comparing the household size estimate to the OLS estimate (obviously if RCS data is being used FE estimates will not be available). The above results show why such an assessment may be erroneous, as variables with a less continuous distribution have more stringent requirements in order to achieve accurate pseudo-panel estimates. This is well exemplified in Column D of Table 3.4a, where although the household size estimate is quite good the other coefficients estimates are rather poor.

Looking at the pseudo-panel estimates in Table 3.4a and 3.4b one can see the efficiency problems associated with having a small number of cohorts. All the pseudo-panel estimates have much larger standard errors than the OLS and FE estimates, which is to be expected, but the standard errors for the estimates in Table 3.4b are also much larger than those in 4a. The reason for this is because the estimates in Table 3.4b all have many more cohorts. However, these estimates are not just better in terms of efficiency but are also less biased and much closer to the

"true" FE values. Thus there does not appear to be any efficiency vs bias trade-off of having larger number of cohorts as highlighted in the literature. Instead increasing the number of cohort appears to be just as important for reducing bias as for increasing efficiency.

This can be clearly seen in Figure 3.3a where we compare the average absolute bias and aggregate absolute bias of each of the 65 regressions against the number of cohorts. Average absolute bias is calculated as the mean across the four regressors of interest (not including the remittance dummy) in each regression, where the absolute bias is the absolute difference between the pseudo-panel estimate and the "true" FE estimate. As this is supposed to be an aggregate measure of the overall regression bias rather than bias of specific regressors, one may worry that certain regressors with higher degrees of bias than others will skew this average measure. For example, one can see from Table 3.5 that the absolute biases of secondary occupation and log remittances are on average over three times larger than for the other two regressors.

To counteract this we construct a second measure called aggregate absolute bias which takes the weighted average across the four regressors but gives three times more weight to household size and log remittances than the other two variables. One sees that with both measures of bias there is a strong negative relationship with the number of cohorts. Interestingly there appears to be a cut-off in both panels around 100 cohorts where estimates become more accurate. This is not just true for these two aggregate measures of bias but also for the bias of individual coefficient estimates, which can be seen in Figure 3.3b. Both the strong negative relationship with cohort numbers and the cut-off at 100 cohorts are present. Also it can be seen that even with low cohort numbers some individual estimates of specific variables can be quite accurate although the other variables may be estimated poorly.

Given how different each of the explanatory variables are, both in terms of their forms and the degree to which they are affected by individual heterogeneity, one may be tempted to reach a general conclusion that having at least 100 cohorts is a useful heuristic when constructing cohorts to ensure pseudo-panel estimates are quite accurate. We do not believe this to be very convincing as we think it is the level of

variation created at the cohort level (which is related to the number of cohorts) which is in fact driving the results. Adding more construction variables to the cohort construction process increases both the number of cohorts and the variation across cohort means. However, it is likely that different construction variables create different amounts of variation for different regressors depending on how strong the relationship between regressor and construction variable is. Thus in the next section we will now try to find which construction variables are the best to use for constructing cohorts, leaving the issues of cohort numbers and the level of variation created for section 6.

Figure 3.3a: Number of Cohorts and Overall Estimation Bias

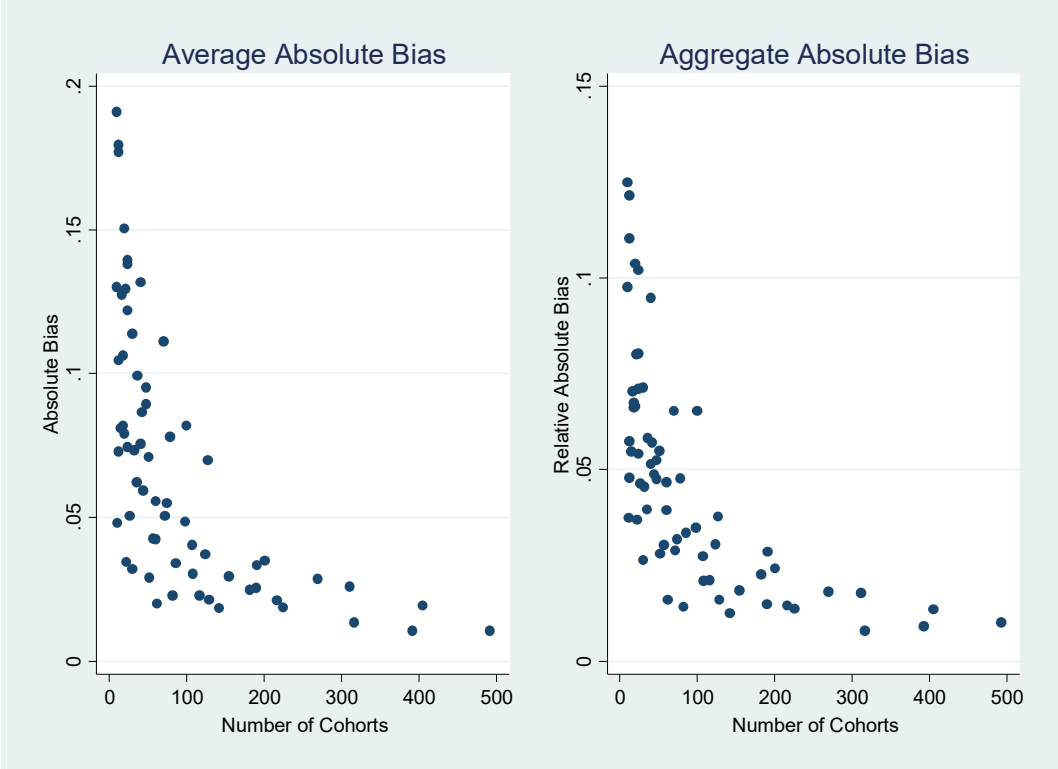
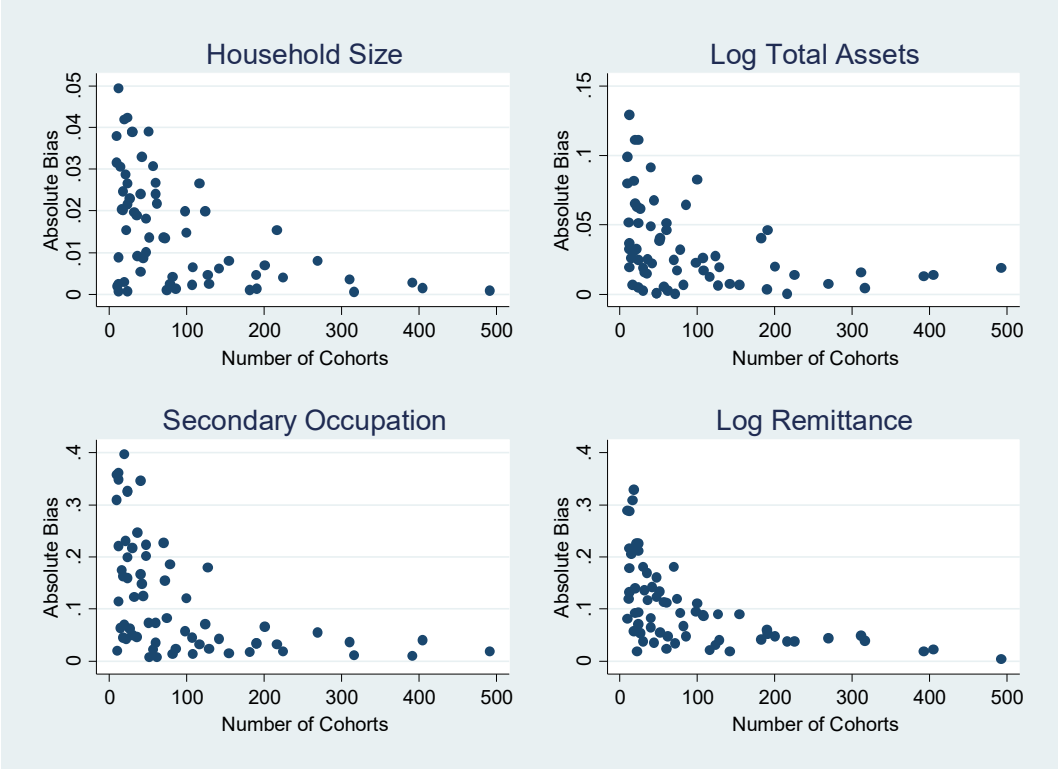


Figure 3.3b: Number of Cohorts and Estimation Bias



5 Optimal Use of Construction Variables

This section considers the five construction variables and assesses their ability to deliver accurate pseudo-panel estimates both individually and in combination with others. The aim is to find some indication of which variable may be better than others to construct cohorts with, not just for our specific dataset but also for other similar datasets, particularly from the LSMS. We find that using age with age bands of no larger than 5 years should be the starting point of any cohort specification. However beyond this it is difficult to make any general recommendations about specific variables or variable combinations to use. Instead what appears more important is the level of variation created amongst the cohorts by a specification.

5.1 Age and the width of its bands

We begin by considering age as a construction variable and the choice of the age bands to use. In Table 3.6 we present a summary of the choice of age bands available, the number of cohorts they group households into and how many of the 65 regressions use the age band. One can immediately see that not using age cohorts can limit cohort construction possibilities if the other construction variables do not create sufficient number of cohorts in order for pseudo-panels to be estimated (with our lower limit being 10 cohorts per time period, thus giving us 40 cohort level in total), hence only six such regressions can be conducted. On the other hand using small age bands, such as 2 year, gives more options and even the choice of just using cohorts based on age. Shorter age bands also allow potentially for more cohort level variation in the data especially if there is a strong relationship between age and the regressors used. The downside of this is that for many such household surveys from developing countries, like the Uganda National Panel Survey, age is likely to be reported with error thus having finer age bands increases the possibility of households being allocated into the wrong cohort, which could potentially be an additional source of bias.

Table 3.6: Age bands

Age Bands	**Number of Cohorts**	**Number of Regressions**
No Age	0	6
17 Year	3	12
10 Year	5	15
5 Year	11	16
2 Year	28	16

Table 3.7: Average Absolute Bias across Age bands

	Obs	**House-hold Size**	**Log Total Assets**	**Secondary Occup-ation**	**Log Remit-tance**	**Average Absolute Bias**	**Aggregate Absolute Bias**
No Age	6	0.022	**0.018**	0.225	0.191	0.114	0.083
17 Year	12	0.014	0.041	0.197	0.153	0.102	0.077
10 Year	15	0.023	0.044	0.158	0.105	0.082	0.066
5 Year	16	0.013	0.028	0.061	0.080	0.046	0.037
2 Year	16	**0.009**	0.029	**0.044**	**0.056**	**0.034**	**0.029**
Overall	65	0.016	0.034	0.119	0.103	0.068	0.053

In Table 3.7 we compare the mean absolute biases for each age band as well as for not using age as part of the cohort construction process. One sees a general pattern across most regressors that bias decreases as finer age bands are used, although there is a noticeable difference in how the secondary occupation and log remittances variables react to finer age bands compared to household size and log total assets. For the former two there is a clear trend of decreasing bias as age is included into the cohort specification and then as shorter bands are used. For household size the two shortest bands do offer the least biased estimates, however the other three categories do not follow such a trend. The results for the log total assets is the most interesting as not using age bands offers the least biased estimates, seeming to go against the trend found in the other three regressors, although the shorter age bands (5 and 2 year) still have less biased estimates then the longer two bands (17 and 10 year). Overall, the results show that 2 year age bands are the best way to construct cohorts although 5 year bands are likely to give acceptable estimates.

In order to explore these finding in more depth and see whether using different age bands has a significant impact on the bias of estimates we conduct a series of regressions which are presented in Table 3.8. We regress the six different measures of bias (four which give the individual bias of each regressor and two aggregate measures) on the different age bands used in the cohort construction process (where not using age as a construction variable is the baseline) and include the four other cohort construction variables as controls. These controls are necessary because as Table 3.6 shows not all the age bands contain an equal number of observations due to needing to exclude any construction specification which produces less than 40 observations in total. Thus the "No Age" category may falsely appear to give less biased estimates purely because the only available construction specifications we have for it are those that use many of the other four other construction variables in combination. For robustness, in Table 3.9 we include not just dummies controlling for the four other construction variables but dummies for all the 15 different ways these four variables may be combined with one another. One could also argue it may be useful to add the number of cohorts created in total as an addition control, however we exclude it due to multicollinearity both with the age bands and the other construction variables.

The coefficients on the age bands in Tables 3.8 and 3.9 show the impact of their inclusion into the cohort construction specification as opposed to age not being included at all in the specification. One can see the decision of whether to include age into the specification or not depends on the bands being used. The 17 year age band does not really appear to make much of a significant difference to the results whereas 10 year bands do show signs of producing better pseudo-panel estimates, particularly for the secondary occupation and log remittance variables. What is clear is that using 5 year or 2 year age bands is very important in getting accurate pseudo-panel estimates and should form the starting point of any cohort construction specification. From Table 3.9 we can see that including 5 or 2 year bands is not only better for reducing bias than including any other single construction variable (as Table 3.8 shows) but also better than any combination of the other four construction variables. Thus there appears to be no substitute for including such bands.

The 10, 5 and 2 year age bands are not only the best bands to use according to our results, but are also the most commonly used bands in the applied literature. It would be thus useful to get a sense of which of the three may be the best to use. As the 5 year is by far the most common we thus compare its performance relative to the other two bands in Table 3.10. We do this by re-estimating all regressions in Tables 3.8 (labelled as “limited”) and 3.9 (labelled as “full”) but this time use 5 year bands as our baseline. Therefore the coefficients on the age bands now show the increase or decrease in bias relative to using 5 year bands, we report these coefficients for 10 and 2 year bands. The 10 year results show that using 5 year bands rather than 10 significantly reduces the bias of all regressors and for both measures of aggregate bias. The size of the reduction in bias is often comparable to adding any of the other four construction variables into the cohort specification, particularly for estimating the coefficients on household size and secondary education. In other words, just using 5 year age bands is often just as good as using 10 year bands combined with any other construction variable. In contrast, moving from 5 to 2 year bands has almost no significant impact on the bias of the estimates apart for some slight improvement in log remittances.

Table 3.8: Relationship between Estimation Bias and Construction variables

	Absolute Bias of:				Average Abs Bias	Aggregate Abs Bias
	Household Size	Log Total Assets	Secondary Occupation	Log Remittance		
17 Year	-0.011	0.010	-0.054	-0.060*	-0.029**	-0.019**
	(0.008)	(0.012)	(0.037)	(0.035)	(0.013)	(0.009)
10 Year	-0.004	0.009	-0.101**	-0.119***	-0.054***	-0.035***
	(0.008)	(0.010)	(0.041)	(0.035)	(0.015)	(0.010)
5 Year	-0.014*	-0.010	-0.204***	-0.148***	-0.094***	-0.067***
	(0.008)	(0.010)	(0.036)	(0.035)	(0.014)	(0.010)
2 Year	-0.019**	-0.009	-0.221***	-0.172***	-0.105***	-0.075***
	(0.008)	(0.010)	(0.036)	(0.035)	(0.014)	(0.009)
Region	-0.011***	-0.006	0.006	-0.053***	-0.016**	-0.014***
	(0.003)	(0.006)	(0.020)	(0.013)	(0.007)	(0.005)
Gender	-0.002	-0.014***	-0.054***	-0.062***	-0.033***	-0.025***
	(0.003)	(0.005)	(0.020)	(0.012)	(0.007)	(0.005)
Location	-0.005*	-0.024***	-0.037*	0.033***	-0.008	-0.010**
	(0.003)	(0.006)	(0.019)	(0.012)	(0.007)	(0.005)
Education	-0.002	-0.035***	-0.080***	-0.042***	-0.040***	-0.033***
	(0.003)	(0.006)	(0.020)	(0.012)	(0.007)	(0.005)
Constant	0.038***	0.078***	0.348***	0.290***	0.188***	0.145***
	(0.010)	(0.012)	(0.046)	(0.039)	(0.018)	(0.013)
Obs	65	65	65	65	65	65
R-squared	0.370	0.570	0.600	0.677	0.726	0.729

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 3.9: Relationship between Estimation Bias and Cohort specifications

	Absolute Bias of:				Average Abs Bias	Aggregate Abs Bias
	Household Size	Log Total Assets	Secondary Occupation	Log Remittance		
17 Year	-0.010	0.006	-0.056*	-0.056	-0.029**	-0.020**
	(0.007)	(0.009)	(0.032)	(0.034)	(0.013)	(0.009)
10 Year	-0.004	0.006	-0.109***	-0.119***	-0.057***	-0.037***
	(0.007)	(0.008)	(0.038)	(0.033)	(0.014)	(0.010)
5 Year	-0.014**	-0.012*	-0.206***	-0.145***	-0.094***	-0.067***
	(0.006)	(0.007)	(0.032)	(0.033)	(0.013)	(0.009)
2 Year	-0.018***	-0.012	-0.223***	-0.169***	-0.106***	-0.075***
	(0.006)	(0.007)	(0.033)	(0.033)	(0.014)	(0.009)
Education	0.001	-0.043***	0.014	0.054**	0.006	-0.003
	(0.012)	(0.005)	(0.054)	(0.022)	(0.011)	(0.009)
Location	0.019	-0.012	0.144**	0.117***	0.067***	0.046***
	(0.012)	(0.010)	(0.059)	(0.035)	(0.021)	(0.015)
Location x Education	0.004	-0.036***	-0.026	0.044	-0.003	-0.007
	(0.012)	(0.012)	(0.053)	(0.042)	(0.012)	(0.010)
Gender	0.004	0.005	0.044	-0.048*	0.001	0.002
	(0.011)	(0.012)	(0.063)	(0.025)	(0.019)	(0.015)
Gender x Education	0.006	-0.042**	-0.041	-0.081***	-0.040***	-0.032***
	(0.012)	(0.017)	(0.036)	(0.026)	(0.010)	(0.009)
Gender x Location	0.007	-0.033***	0.017	0.017	0.002	-0.003
	(0.014)	(0.010)	(0.046)	(0.022)	(0.013)	(0.011)
Gender x Location x Education	-0.005	-0.049***	**-0.093****	-0.039	-0.047***	-0.040***
	(0.012)	(0.009)	(0.042)	(0.026)	(0.013)	(0.010)
Region	-0.010	0.032***	0.144***	-0.008	0.040**	0.030**
	(0.012)	(0.011)	(0.050)	(0.031)	(0.015)	(0.012)

Region x Education	0.003	-0.038***	0.023	-0.046	-0.015	-0.016
	(0.012)	(0.007)	(0.039)	(0.031)	(0.014)	(0.011)
Region x Location	-0.008	-0.029***	-0.008	-0.021	-0.016*	-0.017*
	(0.012)	(0.009)	(0.030)	(0.023)	(0.010)	(0.009)
Region x Location x Education	-0.014	-0.049***	-0.002	-0.034	-0.025**	-0.027***
	(0.012)	(0.007)	(0.040)	(0.022)	(0.011)	(0.009)
Region x Gender	-0.005	0.013	0.075	-0.057**	0.007	0.006
	(0.012)	(0.011)	(0.062)	(0.024)	(0.016)	(0.013)
Region x Gender x Education	0.001	-0.056***	-0.053	**-0.109*** **	**-0.054*** **	-0.045***
	(0.012)	(0.008)	(0.038)	(0.032)	(0.014)	(0.011)
Region x Gender x Location	-0.009	-0.054***	-0.014	-0.006	-0.021***	-0.024***
	(0.012)	(0.008)	(0.034)	(0.026)	(0.007)	(0.007)
Region x Gender x Location x Education	**-0.015**	**-0.059*** **	-0.043	-0.083***	-0.050***	**-0.046*** **
	(0.011)	(0.009)	(0.034)	(0.023)	(0.010)	(0.009)
Constant	0.029**	0.069***	0.255***	0.244***	0.149***	3.817***
	(0.013)	(0.008)	(0.040)	(0.037)	(0.014)	(0.397)
Obs	65	65	65	65	65	65
R-squared	0.566	0.767	0.707	0.776	0.831	0.831

Note: The coefficient estimates in bold are those which signify the largest reduction in bias amongst the 15 cohort specifications. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.

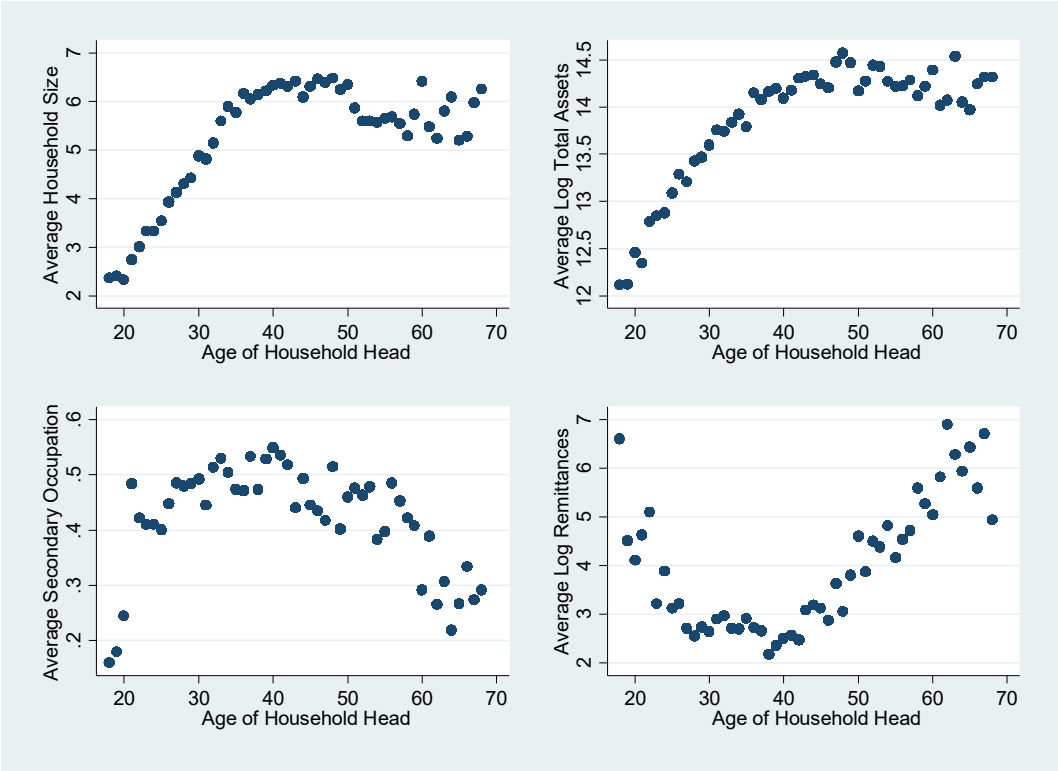
Table 3.10: Re-estimation with 5 Year bands as the baseline comparison

		Absolute bias of:				Average Abs Bias	Aggregate Abs Bias
		House-hold Size	**Log Total Assets**	**Secondary Occupation**	**Log Remit-tance**		
10 Year	Limited	0.010*** (0.004)	0.019** (0.007)	0.103*** (0.032)	0.029* (0.015)	0.040*** (0.011)	0.032*** (0.008)
	Full	0.010*** (0.003)	0.018*** (0.006)	0.096*** (0.030)	0.026** (0.011)	0.037*** (0.009)	0.030*** (0.006)
2 Year	Limited	-0.005 (0.003)	0.000 (0.006)	-0.017 (0.021)	-0.024 (0.014)	-0.011 (0.008)	-0.008 (0.006)
	Full	-0.005 (0.003)	0.000 (0.005)	-0.017 (0.021)	-0.024* (0.013)	-0.011 (0.007)	-0.008 (0.005)

Note: The coefficients in this table refer to the difference in estimation bias between using 5 year bands and using other age bands. The rows described as "Limited" estimate the same model as Table 3.8 and "Full" does the same for Table 3.9. The only difference is the coefficients on the age bands for Tables 3.8 and 3.9 are given in comparison to the baseline of no age cohorts being used, whereas here the baseline is 5 year bands.

The results from Tables 3.7, 3.8 and 3.9 show that the importance of including age cohorts and using smaller age bands differs across different regressors. For secondary occupation and log remittances age is particularly important, even more so than using any other construction variable either on their own or in combination with each other. Age is also quite important for household size, particularly when shorter age bands are used. However for log total assets inclusion of any age band has almost no significant impact on improving the coefficient estimates. We now try and see if there is any ex ante way of anticipating the likely importance of age in estimating coefficients for these four variables. In Figure 3.4 we map the average values for each of the four regressors across the age of the household heads. One can see that household size and log total assets have similar distributions with both initially increasing with age at a diminishing rate before levelling out or even starting to decrease. Similarly, secondary occupation and log remittances also have similar if inverse relationships with age, taking a U or inverse U shape. It should be noted that the relationship seen for log remittances is driven by whether a household receives a remittance rather than the amount of remittances received. The amount received is quite flat and largely unaffected by age however the U shape is present when looking at the relationship between age and if a household gets some remittances or not.

Figure 3.4: Distribution of regressors across age of household head



The U or inverted U shaped distributions of secondary occupation and log remittances suggests a more varied and granular relationship with age than the other two regressors. This may explain why shorter age bands in particular are very important to use for the former two regressors compared to the latter two. However the figures do not help explain why age is important for household size but not log total assets given both regressors have similar age distributions. Therefore distribution figures such as these are at best a rough guide for finding out how important including age may be or what bands to use.

Furthermore, it could be argued that it is not the distribution of the regressors across age which sets the how important age is as a construction variable but instead the continuousness of the regressors themselves. One sees that age and particularly using the shortest age band possible is most important for secondary occupation (a binary variable), followed by log remittances (a variable with both binary and continuous elements), then household size (count) and finally almost insignificant for log total assets (continuous). Therefore as age is the most continuous and only non-

categorical construction variable, it appears its inclusion into cohort construction is of most importance for estimating coefficients of non-continuous variables.

Overall, our analysis indicates that using short age bands (of 5 years or less) should be the starting point of any cohort construction specification. 5 year bands are not just better relative to completely excluding age in the cohort specification but also significantly better than including longer bands such as 10 years. We find little evidence of significant improvements in estimation from using shorter age bands than 5 years, however it may be worth doing so particularly if the regressors of interest are binary or have a strong non-linear relationship with age.

5.2 Categorical Construction Variables

We now discuss and compare the performance of the other four cohort construction variables for obtaining reliable pseudo-panel estimates. We start by looking at each regressor separately to see which construction variable offers the most, if any, improvements to estimates. In Table 3.11 we report the mean absolute bias for the set of 65 regressions which either include each construction variable or exclude it. For household size, for example, we see that including region compared to excluding region has the largest reduction in bias than for any of the other three categorical constructional variables. Inclusion of location has a very small impact on reducing the bias, however for gender and education there is little difference in the mean absolute biases between specifications that include or exclude them. These results are consistent with the findings in Table 3.8 where we can see if the construction variables significantly affect the bias after controlling for the different age bands. One sees that including region significantly reduces the bias, location also reduces bias but the significance level is lower, and gender and education are insignificant.

Table 3.11: Absolute bias across constructional variables

	Obs	Household Size	Log Total Assets	Secondary Occupation	Log Remittance	Average Abs Bias	Aggregate Abs Bias
No Region	28	**0.021**	0.034	0.095	0.120	0.067	0.054
Region	37	**0.011**	0.033	0.138	0.091	0.068	0.053
No Gender	30	0.016	0.040	0.140	**0.129**	0.081	0.063
Gender	35	0.015	0.028	0.102	**0.082**	0.057	0.045
No Location	30	0.017	0.045	0.130	0.078	0.068	0.056
Location	35	0.014	0.023	0.110	0.125	0.068	0.052
No Education	29	0.016	**0.052**	**0.150**	0.115	**0.083**	**0.067**
Education	36	0.015	**0.019**	**0.095**	0.093	**0.056**	**0.043**
Overall	65	0.016	0.034	0.119	0.103	0.068	0.053

For log total assets, the results from both Tables 3.8 and 3.11 show a consistent picture with education leading to the largest reduction in bias, followed by location and gender which also significantly reduces bias and finally with region which has no significant impact. Similar results apply for secondary occupation with region again being insignificant, education causing the largest reduction in bias, gender and location leading to significant reductions although location is less significant than it was for log total assets. For log remittances gender has the largest effect on reducing bias, followed by region and then education all of which have highly significant effects. Interesting location has a positive and significant effect on the bias, being the only instance of where the inclusion of a construction variable significantly worsens estimates.

The above analysis of individual regressor biases show that it is not obvious that one construction variable is superior to all others for inclusion into the cohort specification. None of the construction variables significantly reduce bias for all four regressors. In fact, location significantly reduces the bias for three regressors but significantly increases it for the fourth. Education and gender seem to be in general the best variables to include with both being highly significant for three of the four regressors and having the largest impact for reducing bias in the two aggregate measure regressions. Region is also a useful variable to include having relatively large and significant impact on bias reduction in four of the six regressions.

Thus one can see the importance of each construction variable differs across the regressors, so it would be useful to see if there is any way of predicting which construction variables would be best to use for specific regressors. It is worth noting that unlike with age the continuousness of the regressor makes no difference. For example, region is insignificant for log total assets (the most continuous variable) and secondary occupation (a binary variable) but significant for the other two regressors. Similar results apply for the other three construction variables. Instead it may prove useful to see how much regressors vary across each of the categories of the construction variables. In Table 3.12a we present the mean and standard deviation of household size across the different subcategories of each construction variable. One would expect that the more variation there is in the means across

subcategories the more variation there would be at the cohort level, hence the better pseudo-panel estimates are likely to be. This appears to hold for region and education. Household size varies the most across region than any other construction variable, which may explain why region causes the largest and most significant fall in bias for household size estimates. With education one sees that the means only really differ between households with heads with no education and those with either primary or post-primary education. Even then the difference is smaller than that observed across subcategories of the other variables, thus potentially explaining why education is relatively poor for household size. However, with gender and location this pattern does not hold. We see that location is the only one of the two to have a significant negative effect on the bias even though it is gender which has the larger degree of variation across the means of its subcategories.

For log total assets, we see that a largely similar pattern also holds. We know that education and then location are the best construction variables to use for reducing bias, and Table 3.12b also shows that these two variables (in the same order) produces the greatest variation across the means in their subcategories. For region and gender this does not hold as much, as it is region that produces more variation in the means even though only gender is found to significantly reduce the bias of the estimates. For secondary occupation, the pattern appears to breakdown completely with education and region being found to have the largest and most significant impact on bias reduction even though Table 3.12c shows the two produce by far the lowest variation across subcategory means. Location, which creates the greatest difference in means, has only a small significance in reducing bias and region, which has a higher degree of variation than education or gender, is completely insignificant. The pattern is also quite weak for log remittances, where gender followed by region and then education have the largest and most significant impact on bias reduction whereas location actually significantly increases the bias. One may have been able to predict gender having the largest effect as Table 3.12d shows there is a very large difference in the proportion of households receiving a remittance based on whether the head is male or female, although the actual amounts remitted are not as different as those found amongst subcategories of other construction variables. For region and education, there is some variation both in the proportions of households that received

remittances and the amounts remitted to anticipate that these two variables will be useful for estimating log remittance coefficients. However, for location it is hard to predict the significant positive effect that it would have on the bias given that there is quite a strong amount of variation in the amounts remitted even if the proportions that received remittances are not too different. In particular, comparing the subcategory means for location and education, it is hard to explain why one has a significant positive effect on the bias and the other a significant negative effect.

Table 3.12a: Distribution of Household Size across cohort construction variables

Household Size					
	Region				
	Central 31.1%	Eastern 23.1%	Northern 24.5%	Western 21.2%	Total
Mean St Dev	5.03 2.95	6.13 3.03	5.56 2.47	5.52 2.71	5.52 2.84
	Gender		**Location**		
	Male 72.0%	Female 28.0%	Rural 75.7%	Urban 24.3%	Total
Mean Variance	5.81 2.88	4.79 2.57	5.69 2.79	5.01 2.92	5.52 2.84
	Education				
	None 15.2%	Primary 53.7%	Post-Primary 31.1%		Total
Mean Variance	5.11 2.52	5.68 2.85	5.45 2.93		5.52 2.84

Table 3.12b: Distribution of Log Total Assets across cohort construction variables

Log Total Assets					
	Region				
	Central 31.1%	Eastern 23.1%	Northern 24.5%	Western 21.2%	Total
Mean Variance	14.31 2.13	13.96 1.73	13.51 1.86	14.00 1.68	13.97 1.91
	Gender		**Location**		
	Male 72.0%	Female 28.0%	Rural 75.7%	Urban 24.3%	Total
Mean Variance	14.08 1.86	13.66 2.00	13.83 1.74	14.38 2.31	13.97 1.91
	Education				
	None 15.2%	Primary 53.7%	Post-Primary 31.1%		Total
Mean Variance	13.08 1.94	13.74 1.71	14.79 1.91		13.97 1.91

Table 3.12c: Distribution of Secondary Occupation across cohort construction variables

Secondary Occupation					
	Region				
	Central 31.1%	Eastern 23.1%	Northern 24.5%	Western 21.2%	Total
Mean	0.38	0.44	0.57	0.44	0.45
Variance	0.49	0.50	0.50	0.50	0.50
	Gender		**Location**		
	Male 72.0%	Female 28.0%	Rural 75.7%	Urban 24.3%	Total
Mean	0.46	0.43	0.50	0.32	0.45
Variance	0.50	0.50	0.50	0.47	0.50
	Education				
	None 15.2%	Primary 53.7%	Post-Primary 31.1%		Total
Mean	0.45	0.47	0.43		0.45
Variance	0.50	0.50	0.50		0.50

Table 3.12d: Distribution of Remittances across cohort construction variables

Remittances						
Region						
		Central 31.1%	Eastern 23.1%	Northern 24.5%	Western 21.2%	Total
Remittance Dummy	Mean	0.34	0.31	0.32	0.18	0.29
	Variance	0.47	0.46	0.47	0.39	0.46
Log Remittance	Mean	12.53	11.76	11.44	11.41	11.91
	Variance	1.73	1.58	1.71	1.64	1.75
		Gender		**Location**		
		Male 72.0%	Female 28.0%	Rural 75.7%	Urban 24.3%	Total
Remittance Dummy	Mean	0.20	0.53	0.28	0.33	0.29
	Variance	0.40	0.50	0.45	0.47	0.46
Log Remittance	Mean	11.54	12.27	11.62	12.66	11.91
	Variance	1.75	1.67	1.65	1.78	1.75
		Education				
		None 15.2%	Primary 53.7%	Post-Primary 31.1%		Total
Remittance Dummy	Mean	0.36	0.27	0.30		0.29
	Variance	0.48	0.44	0.46		0.46
Log Remittance	Mean	11.38	11.62	12.66		11.91
	Variance	1.50	1.70	1.71		1.75

This method of using the variation in the means across subcategories of construction variables to predict how important their inclusion into the cohort specification for a regressor works quite well for log total assets and to household size, less so for remittances and not very well for secondary occupation. One could think that it may be the continuousness of the regressor affects how well this method works (poorly for binary variables and well for continuous). However, it should also be noted that the variables for which this method does not work are also the variables for which using age cohorts, particularly one with short bands, become very important, in fact far more important than including any other construction variable. Thus it is likely that how well a categorical construction variable performs for estimating these regressors depends how these construction variables interact with the age bands and the amount of variation the interaction creates. Furthermore, how the categorical construction variables interact with each other is also of crucial importance. One can see from Table 3.9 that the best specifications (in terms of having the largest and most significant negative effect on the bias) are those that include all four construction variables or at least three of the four. All this makes it quite hard to say that a specific construction variable should in general be favoured or prioritised. Although there is some evidence that education and gender are better than region or location, what appears more important is the number of construction variables included in total and how they interact with each other. Therefore it may be better to set the question up differently, where one starts off with a full specification of all the construction variables and then, if needed, exclude the variables whose exclusion is likely to be the least detrimental rather than trying to pick the ones which may be most beneficial. The reason such exclusions may be necessary is to ensure cell sizes are sufficiently large to deal with measurement error.

We present an example of how this can be done below in Table 3.13. In the first two rows we present the “true” fixed effects coefficients and the pseudo-panel estimates for when the cohorts are constructed using 2 year age bands combined with all four of the other construction variables. One can see the pseudo-panel estimates from such a specification are all close to the “true” values for all regressors. However the cell sizes are very small (less than 5 on average) and such a specification will suffer from a great deal of measurement error if repeated cross-sections rather than panel

data was used. To increase cell size we remove two variables from the specification based on the amount of variation they produce for each regressor as shown in Tables 3.12a-d. So for household size we remove location and education as these have the lowest variation in the means across their subcategories, presently the results in the third row. For log total assets we remove gender and location (fourth row), gender and education are removed for secondary occupation (fifth row) and finally for log remittances we remove region and location (final row). The decision was slightly harder for log remittances because education and region have similar variations. Education was included as it has greater variation in the amount remitted which complements the inclusion of gender, which has less variation in the amount remitted but a great deal in the proportion of households that receives any remittance. The shaded cells connect the regressor estimates to the specification chosen for that regressor. One sees that these four specifications with two construction variables removed are all worse at estimating the whole model than if we included all variables in the specification. However, they do perform well in estimating the specific regressor the specification is chosen for, in some cases being closer to the "true" values than the full specification estimates. This shows that if it is not possible to use a full specification due to cell size problems, one can instead estimate with multiple specifications of fewer construction variables where each specification is targeted for a specific regressor using information on their variation as shown in Tables 2.12a-d. Thus Tables 2.12a-d may not be useful in showing which regressors should be included from the specification, but can help in identifying which ones to potentially exclude.

Table 2.13: Pseudo-panel estimates from different cohort specifications

House-hold Size	Log Total Assets	Secondary Occupation	Log Remit-tance	Cell Size	Number of Cohorts	Cohort Specification
-0.090	**0.100**	**0.048**	**0.026**	**-**	**-**	"True" FE coefficients
-0.089	0.118	0.067	0.030	4.9	435	2 year, all variables
-0.091	0.146	0.082	0.078	13.4	183	2 year, region, gender
-0.082	0.107	0.103	0.070	9.4	259	2 year, region, education
-0.091	0.140	0.064	0.067	14.1	173	2 year, region, location
-0.084	0.092	0.005	0.043	18.1	136	2 year, gender, education

To summarise our findings thus far; we recommend cohort construction specifications should use age bands of no more than 5 years. Shorter bands may be required if regressors have a non-continuous distribution (particularly if they are binary) and if their distribution across age is highly non-linear (U or inverse U shaped like in the bottom panels of Figure 3.4). For the other construction variables, it is hard to identify any specific variable which works best, with some weak evidence favouring education and gender. Instead the best approach is to start off using as many construction variables as possible and then pick which ones to omit based on how much the means of regressors vary across their subcategories.

6 Assessing Cohort Construction

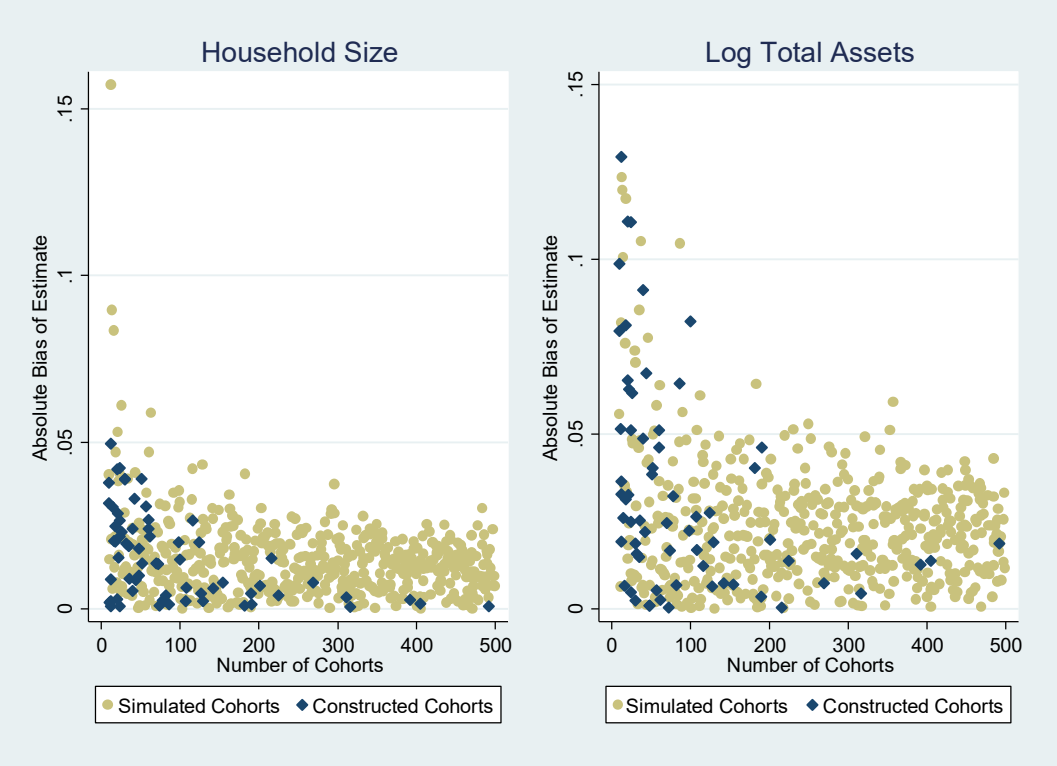
The last section demonstrated the importance of the overall cohort specification rather than individual construction variables for deriving accurate pseudo-panel estimates. In this section we derive some measures that can be used to assess whether a specification is likely to deliver reliable estimates. We first propose a new measure, called AWAR, which considers three important sources of variation in the cohort level data. Then, noting the similarity between cohort construction and using instrumental variables, we compare AWAR to four other measures based on tests of instrument validity. We find that when AWAR is above 0.5 the cohort specification delivers accurate pseudo-panel estimates that are similar to panel FE estimates. Furthermore, while some of the IV based measures can be useful in assessing cohort construction we argue that AWAR is the preferable option.

6.1 Deriving AWAR

Looking at Figures 3.3a and 3.3b, which show the relationship between the number of cohorts a specification generates and biases of the estimates, one can see a general trend of estimates improving as the number of cohorts increase. Nevertheless, even at low numbers reliable estimates are still possible but there is a high possibility of estimating erroneous coefficients. There does appear to be a cut off around the 100

to 150 mark where biases become much smaller. However this in itself may not be a useful guideline to follow for cohort construction because increasing the number of cohorts should in theory just improve the efficiency of estimates and if anything increase the bias given cell sizes (the number of households per cohort) are decreasing, thus causing greater measurement error. Instead we believe the relationship seen in in the Figures is mainly driven by the level of variation created at the cohort level which impacts directly on the bias of the estimates but is also related to the number of cohorts created.

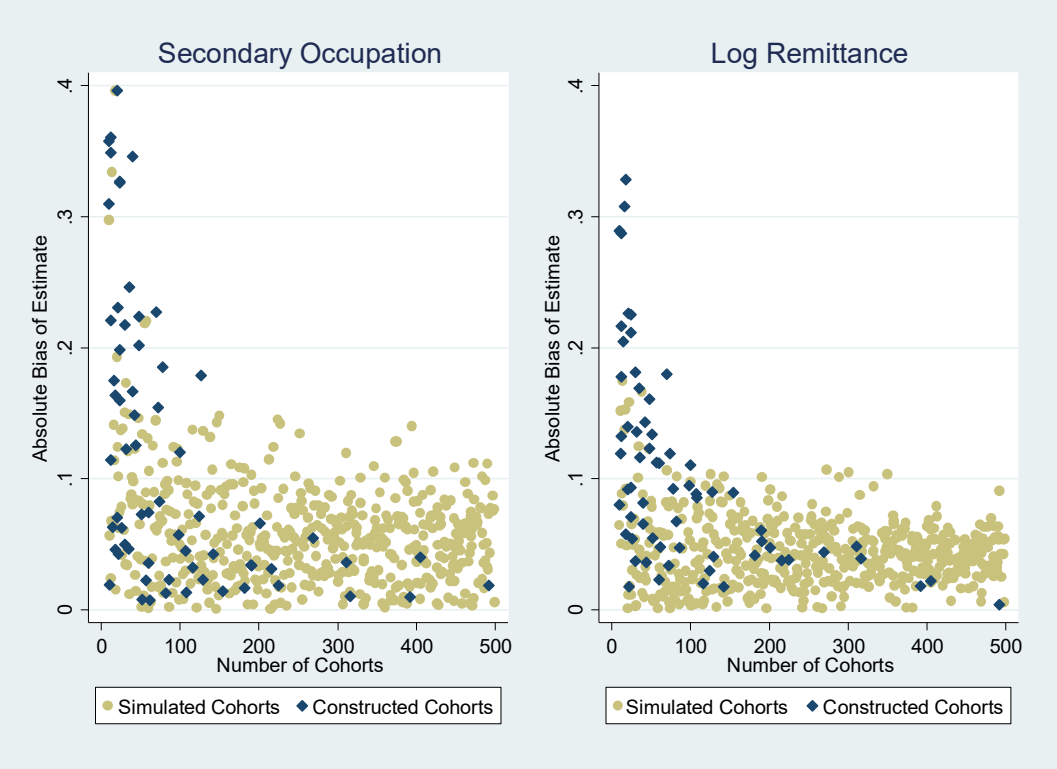
Figure 3.5a: Simulated and Constructed Cohort Estimates



Evidence for this can be seen in Figures 3.5a and 3.5b where alongside the 65 constructed cohort specification we also plot results from cohorts which were created at random, which we call "simulated cohorts". For these simulated cohorts, we randomly assign households into a cohort for all time periods and vary the number of cohorts created from 10 to 500. Thus we have an additional 490 estimates from these simulated cohorts. The simulated cohorts allow for cohort numbers to be increased without having much effect on the cohort level variation. One can see that even with

nearly 500 cohorts the pseudo-panel estimates are still not very reliable. Thus one cannot devise general rules of thumb for cohort construction based on the number of cohorts. This is similar to the point made above where a general recommendation regarding cell sizes also failed to deliver accurate estimates. This is because the recommendation of 100-200 cell sizes was given with the caveat of a specific degree of cohort level variation in the original Verbeek and Nijman (1993) paper which has since often been ignored.

Figure 3.5b: Simulated and Constructed Cohort Estimates



This highlights that the key to understanding how likely a specification will deliver accurate estimates is to assess the variation created. We identify three different types of variation important for the estimation of reliable coefficients. For the first two we follow Verbeek and Nijman (1992) who found two factors were important in reducing bias: the true variance across cohort means (w_1) and the measurement error variance of the observed cohort means (w_2), which are represented above in equations (4) and (5). Similar concepts can be found elsewhere in the literature often

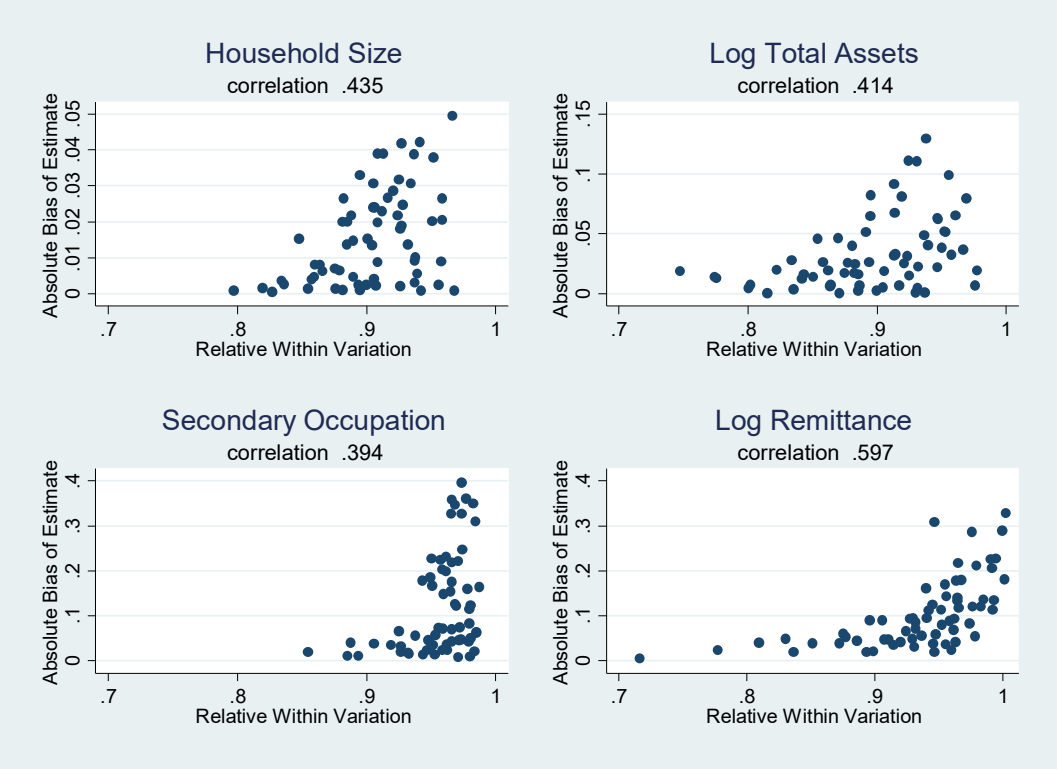
referred to as inter-cohort heterogeneity (for w_1) and intra-cohort homogeneity (for w_2) as the measurement error can be reduced by increasing homogeneity and grouping similar individuals together. Verbeek and Nijman show having w_1 as large as possible relative to σ_v^2 (variance of individual x_{it} observations within each cohort) reduces the cell sizes required to minimise the bias. They also give an example for when cell size is 100, showing that a ratio of $w_1 / \sigma_v^2 = 0.5$ is sufficient to produce accurate estimates. Given that x_{ct}^* (the true cohort population means) cannot be observed for our data we begin by trying to identify proxies for w_1 and w_2.

For the measurement error variance (w_2) we produce a measure which first calculates the standard deviation for households within each group (a group is a cohort in a specific time period) and then take the average across all groups weighted by the square-root of the cell size for each regressor. We call this measure within variation (or $\hat{\sigma}_v$) and we predict this will be positively related with estimation bias as higher variance means cohorts contain a less homogenous set of individuals, therefore cohort means are likely to be measured with greater error. It should be noted that within variation only captures measurement error around the observed cohort sample means, which themselves are likely to be measured with error, rather than the true cohort population means (which are required for calculating w_2 but are unobserved). We then rescale the within variation by dividing it by the overall standard deviation at the household level. This produces a measure we define as the relative within variation which essentially captures the homogeneity of households within a group compared to the overall dataset.

Figure 3.6 shows the relation between relative within variation and estimation bias for our four regressors. As predicted a positive correlation exists and although it is in most cases quite moderate one can see that when the variation is at its lowest the bias of the estimates is also likely to be low. There also appears to be thresholds below which the probability of getting reliable estimates are much greater. The thresholds are around the 0.9 mark although there is some variation depending on the regressor, particularly its continuousness. The regressors which are the most continuous (log

total assets and household size) require the strictest threshold of around 0.8 to 0.85, whereas log of remittances and secondary occupation have thresholds at 0.9. The threshold levels appear to be quite sensitive with a 0.9 threshold failing to reject very poor specifications for the total assets and household size variables, whereas a 0.85 threshold for secondary occupation would eliminate all specifications.

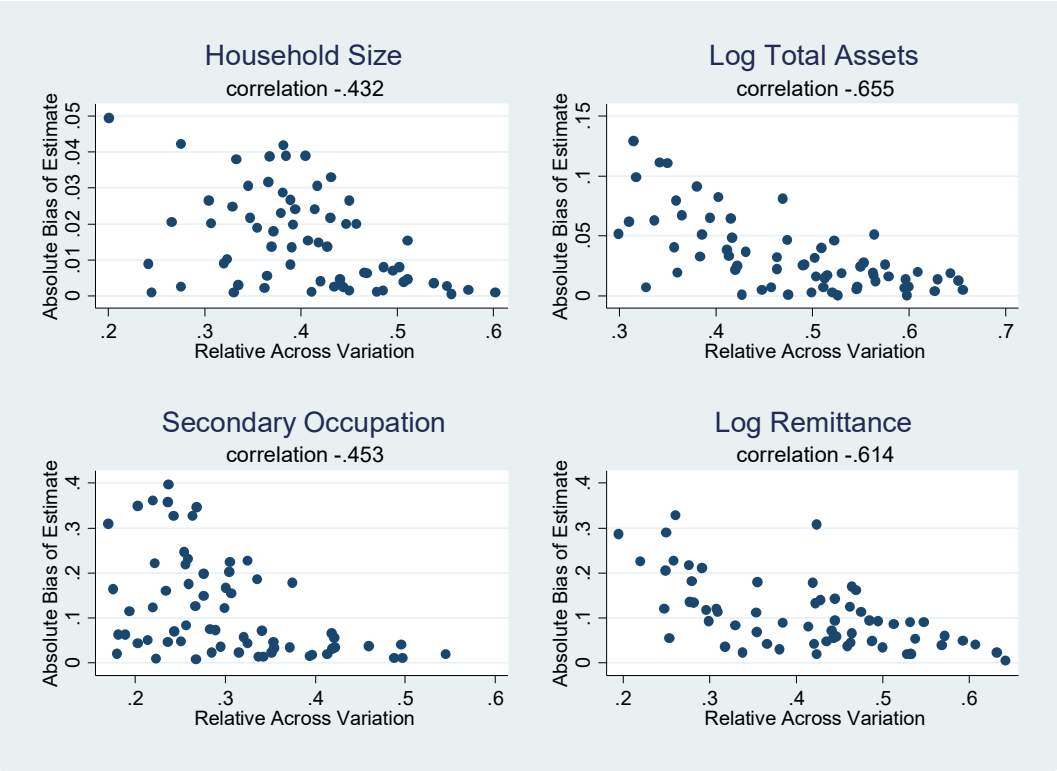
Figure 3.6: Relative Within Variation and Estimation Bias



Our proxy for inter-cohort heterogeneity (w_1), which we call across variation (or $\hat{\sigma}_x$), is derived by calculating the standard deviation across the cohort means weighted by the square-root of the cell size. This measure imperfectly captures w_1 as it calculates the variation around observed cohort sample means not the true population means, thus it may pick up some variation that is due to measurement error and not purely variation across the true cohort means. We then rescale the measure by dividing it by the overall standard deviation at the household level as before to get the relative across variation.

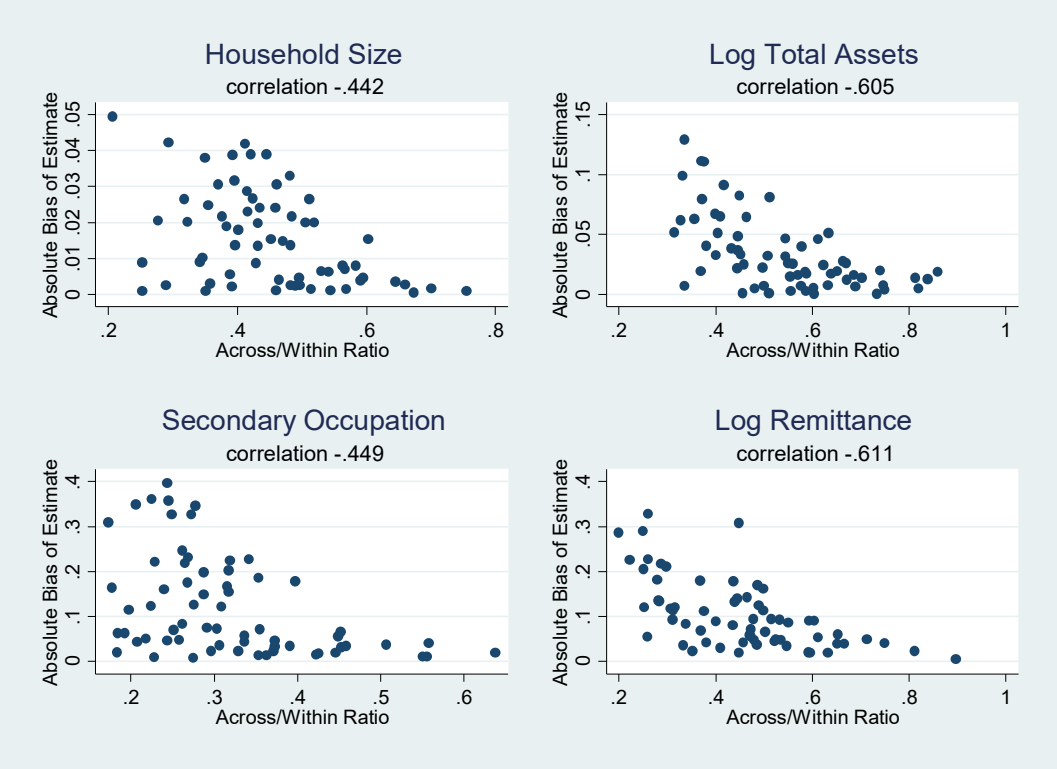
Figure 3.7 shows the relationship between this measure and the absolute biases of each regressor. A negative relationship can be seen as stipulated in Verbeek and Nijman (1992) where a higher degree of variation leads to more accurate estimates and the correlations are higher than observed for the within variation. One way of interpreting this relationship is to see the measure as a ratio of the variation at the cohort level compared to the household level. If the cohort level has too little variation comparatively, it is likely that there will be aggregation bias and hence the estimates are less comparable to the household level estimates. It is not that estimates are necessarily biased, more so that they are only suited for a more aggregate level of analysis rather than at the household level. However as this study is motivated towards the latter, thus we want estimates to be as close to the household level estimates as possible. In terms of useful thresholds for finding a sufficient level of aggregation, there seems to be a rough threshold between 0.4 and 0.5. Like in Figure 3.6, a stricter threshold (now the higher one of 0.5) is required for variables with the most continuous distribution and a looser threshold of 0.4 for the more discrete.

Figure 3.7: Relative Across Variation and Estimation Bias



Verbeek and Nijman (1992) states that in order to reduce bias what is required is to increase the true variance across cohort population means (w_1) relative to the within variation (σ_v^2) and gives a ratio of 0.5 as being sufficient for reasonable unbiased estimates. We calculate this ratio using our proxies for both terms ($\hat{\sigma}_x / \hat{\sigma}_v$) and show the relationship with estimation bias in Figure 3.8. An Across/Within ratio of 0.5 may be a reasonable threshold for the secondary occupation variable, however other variables need a much stricter and higher threshold, potentially as high as 0.8 for the log of total assets. One may notice the ratio denoted by Verbeek and Nijman using the variances whereas our proxies use standard deviations, therefore a threshold of 0.7 (roughly the square-root of 0.5) is more suitable. However even this threshold fails to work across all four regressors, being too strict for secondary occupation and household size. Thus, as before, defining a single threshold still appears difficult.

Figure 3.8: Across to Within Variation and Estimation Error



We believe the reason Verbeek and Nijman's Across-to-Within ratio was unable to deliver clear thresholds for assessing whether the cohort construction can produce

accurate estimates is because it ignores a third important source of variation in the data; the variation over time in the cohort means. Devereux (2007a, 2007b) shows that if there is insufficient variation across time relative to variation across cohorts in the same time period this may exacerbate measurement error problems and lead to serious small sample biases. In order to incorporate time variation we assume the cohort level explanatory variables follow an AR[1] specification as in equation 11 below:

$$x^*_{ct} = \rho x^*_{c(t-1)} + e_{ct} \qquad (11)$$

Where x^*_{ct} represents the true cohort population means and e_{ct} is IID $N(0, \sigma^2_e)$. Consequently the variance of the cohort population means (w_1) is:

$$w_1 = \frac{\sigma^2_e}{1 - \rho^2} \qquad (12)$$

Verbeek and Nijman implicitly assume ρ=0 and therefore $w_1 = \sigma^2_e$. Thus it is also assumed that the overall level of variation in the cohort population means (w_1) is equivalent to the genuine level of variation across cohort observations (σ^2_e). However once autocorrelation is introduced w_1 increases without there being any additional genuine variation across cohorts. Alternatively one can think of it as w_1 remaining unchanged but there is a fall in the level of genuine variation across cohorts. Thus without correcting for autocorrelation w_1 and our proxy for it ($\hat{\sigma}_x$) is likely to overestimate the variation across the cohort level observations. This means the relative across variations calculated in Figure 3.7 are biased upwards as the cohort observations may be autocorrelated but the household level observations cannot be as they are independent across time. Therefore the ratio of interest should instead be σ^2_e relative to σ^2_v which can be calculated as shown in equation 13:

$$\frac{\sigma^2_e}{\sigma^2_v} = \frac{w_1(1-\rho^2)}{\sigma^2_v} \qquad (13)$$

Equation 14 shows how such a measure can be constructed using the proxies for the three types of variation identified. We calculate $\hat{\sigma}_x$ and $\hat{\sigma}_v$ as before, $\hat{\rho}$ is calculated

by regressing the explanatory variables on their first lag and using the square-root of the cell sizes as weight. As the proxies use standard deviations we thus take the square-root of $1-\hat{\rho}^2$ for consistency. This results in what we call the Across-to-Within Autocorrelation Adjusted Ratio, or AWAR for short.

$$\text{AWAR} = \frac{\hat{\sigma}_x (1-\hat{\rho}^2)^{1/2}}{\hat{\sigma}_v}$$

Figure 3.9: AWAR and Estimation Bias



The intuition behind the ratio is that higher genuine across variation (where genuine means discounting the effect of autocorrelation on the cohort level variation) reduces aggregation bias in the cohort level data. This bias is worsened if there is higher within variation because it indicates households within a cohort are more heterogeneous meaning cohort averages are less representative, causing measurement error bias. Therefore AWAR is expected to have a negative relationship with estimation bias. In Figure 3.9 we plot the AWAR values for each cohort construction specification against the bias for each regressor of interest. Now

a clear threshold becomes evident, particularly due to the non-linear relationship between bias and AWAR, at 0.5 for all regressors above which estimates are quite accurate. This is not to say that below this threshold all individual estimates for specific regressors will be inaccurate, as the figure shows accurate estimates can be achieved even for very low AWAR values. Instead the threshold ensures confidence that any estimate produced will not suffer from substantial bias caused by the cohort aggregation process.

6.2 AWAR and IV-based Alternatives

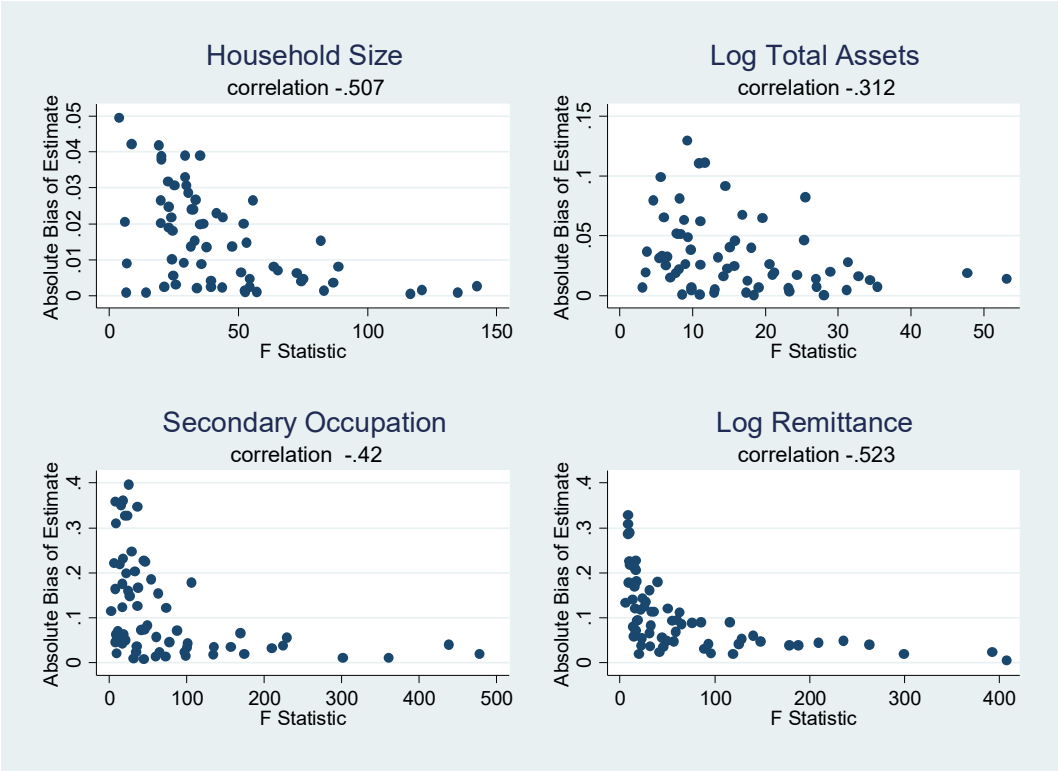
Before assessing how general the results seen in Figure 3.9 are we consider some alternative way to assess whether the cohort level observations reflect the data at the individual level by expanding on the view of Moffitt (1993) where cohort aggregation is seen as similar to using instrumental variables. In fact if one estimates the model at the individual level but includes cohort dummies while using the cohort means as instruments [as shown in Equation (15)] then the exact same estimates are achieved as when using pseudo-panel estimation, although the standard errors differ. Therefore one can use some of the tests for instrument validity to assess cohort construction as the cohort means will be weak instruments if they are not sufficiently correlated with the household level values.

$$y_{it} = x'_{it}\beta + f_c + f_t + \varepsilon_{it} \qquad \text{using } \bar{x}_{ct} \text{ as instruments for } x_{it} \qquad (15)$$

One of the most common ways to measure validity is to use the F-statistic from the first stage IV regression with a general rule of thumb being that a value of more than 10 indicates that instruments are valid. Figure 3.10 reports the F-statistic and absolute bias for each of the four regressors and a pattern clearly emerges where estimates improve as the F-statistic increases. However it is evident that a threshold value of 10 is not a useful heuristic to assess cohort construction. In fact it is difficult to identify a threshold common to all regressors; with log total assets appearing to have one at 30, which would be far too lax for any of the other regressors, and secondary occupation requiring one closer to 100, which would be too stringent for others. Devereux (2007b) reached a similar conclusion regarding the F-statistic as a

possible indicator of bias, stating it was of limited value because it was unclear how large it needed to be. In contrast, comparing Figures 3.9 and 3.10, it is clear AWAR does have an identifiable threshold as well as stronger correlations with the absolute biases, making it a better measure for assessing cohort construction than the F-statistic.

Figure 3.10: F-Statistic and Estimation Bias



The R-squared from the first stage IV regressions can also be used to ascertain the strength of the relationship between the cohort and household level observations. This is because the first stage consists of regressing the endogenous variables [i.e. the household level variables (x_{it})] on the exogenous variables and instruments [i.e. the cohort means ($\bar{x}_{ct}$)]. Figure 3.11 shows that much like the F-statistic, there is a strong relationship with estimation bias (with correlations being similar to those of AWAR) but a threshold R-squared value consistent for all regressors is hard to identify; potentially being anywhere between 0.2 and 0.4 with either end of the spectrum being either too lax or too stringent for at least one of the regressors.

Figure 3.11: First-stage Regression R-squared and Estimation Bias

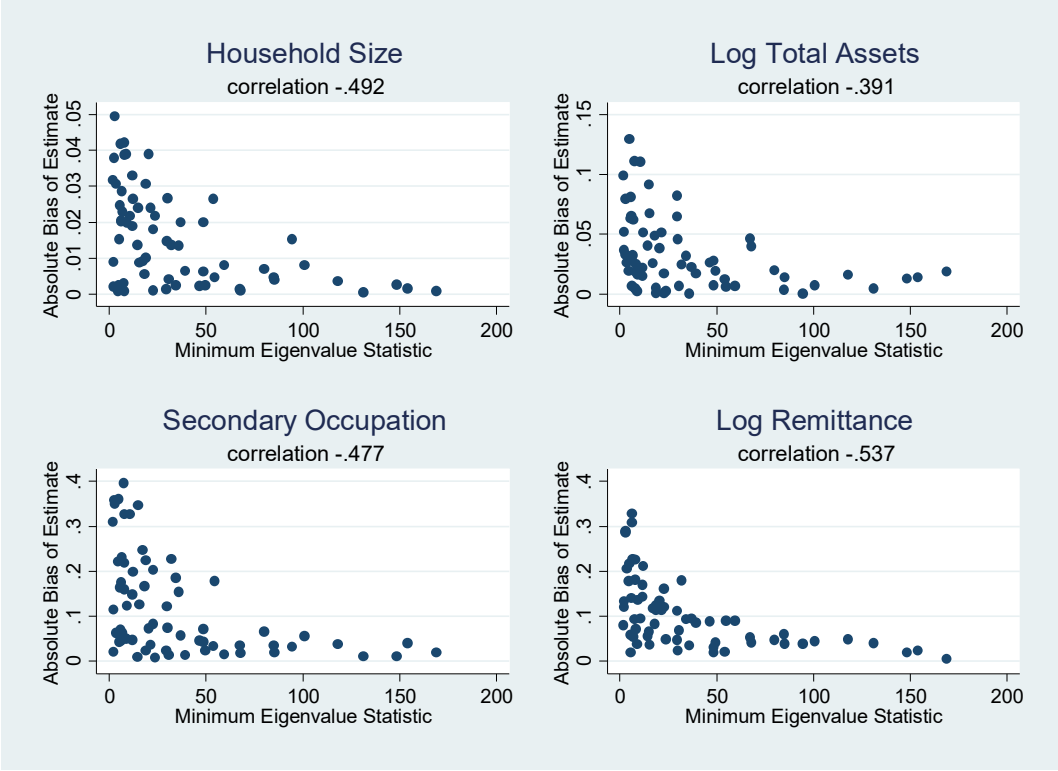


Figure 3.12: Shea's Partial R-squared and Estimation Bias



A related but alternative measure of instrument weakness is Shea's partial R-square, which measures the correlation between the endogenous variables and instruments after partialling out the effect of exogenous variables. Although there are no rules of thumb or formal distribution theory that gives a threshold value required for ensuring instrument validity, Figure 3.12 indicates a threshold exists for the similar purpose of cohort construction. For all regressors, once the R-squared is above 0.07 pseudo-panel estimates are quite accurate compared to the "true" household level estimates.

Figure 3.13: Minimum Eigenvalue Statistic and Estimation Bias



The final measure of weak identification considered is the Minimum Eigenvalue Statistic (henceforth MES) for which a formal test exists using the critical values tabulated by Stock and Yogo (2005). The MES is constructed based on the correlation between the set of endogenous variables and instrumental variables, rather than for each endogenous variable separately. Therefore it is a combined test of weak identification for all endogenous variables and unlike the measures considered thus far it cannot be used to test the validity of one specific variable. In

Figure 3.13 we compare this single aggregate measure to the bias of each of the four regressors. The critical values relevant for our estimates would at most be 20 or 30 according to the Stock and Yogo tables, however the figure shows this would be too low with heavily biased estimates being produced at such vales. Instead, values between 50 and 100 give reasonably accurate estimates for most regressors. Once the MES is above 100 then estimates for all regressors become accurate without the threshold being overly stringent as several cohort construction specifications are able to satisfy it. One point worth noting is that in order to calculate the MES the IV regression cannot be weighted whereas for consistency with the pseudo-panel estimates all other IV regressions conducted thus far have been weighted by negative of the square-root of cohort size.

Table 3.14: Correlation between AWAR and weak instrument measures

	Household Size	**Log Total Assets**	**Secondary Occupation**	**Log Remittance**
Shea's R-squared	0.96	0.98	0.99	0.94
MES	0.94	0.97	0.98	0.94
First Stage R-squared	0.98	0.92	0.99	0.74
F-statistic	0.95	0.89	0.94	0.91

Evidently cohort aggregation and instrument validity are related processes with all four validity measures displaying a strong non-linear relationship with the bias arising from aggregation. The finding that the F-statistic and MES values required for reliable pseudo-panel estimates exceed those required for instrument validity demonstrates cohort aggregation is a more exerting and stringent process, further evidence of why aggregation needs to be conducted with more caution then is presently done in empirical applications. The AWAR measure proposed for assessing cohort construction we believe is a more useful statistic than the other four validity measures considered, although as Table 3.14 shows it is very highly correlated with all four particularly Shea's R-squared. The F-statistic and first stage R-squared have no consistent threshold across all four regressors thus it is a hard to identify some heuristic to assess the aggregation process. As the four regressors all differ in both the degree they are affected by unobserved heterogeneity and their

form (how discrete or continuous they are), this means that any threshold consistent with all four variables is quite a general result.

Compared to AWAR and Shea's R-squared, the MES has a less identifiable threshold lying between 50 and 100. This is likely due to the fact that it is a measure of validity for the whole set of endogenous variables (of all four regressors in our case) thus it is less suited to identifying thresholds for a specific regressor. This is an important drawback as cohort aggregation is likely to have different effects on each regressor depending on the relationship between the cohort construction variable and the individual regressors. Thus the MES figure may hide the fact that cohort aggregation works very well for some regressors but not for others. Both AWAR and Shea's R-squared do not suffer from this drawback and both have quite clear and general thresholds. One could argue that AWAR is slightly better as the correlation with aggregation bias is stronger and the thresholds for secondary occupation and log remittances are more identifiable (with Shea's R-squared one could make the case to have a threshold closer to 0.05). Furthermore AWAR can be deconstructed to the three underlying types of variation discussed above, allowing one to identify the source of bias and potentially pick alternative construction variables in the cohort specification which could address that specific variation.

In terms of applications of these tests we recommend the thresholds be applied to all regressors of interest rather than taking an average score across regressors and applying the threshold to that average. This is necessary as some variables may score very highly on the measures (particularly for AWAR if time correlation is very low) which would inflate the average scores and potentially hide the lack of variation in other variables. This is particularly important given the nature of the non-linear relationships (often L-shaped) between bias and the measure scores indicating that relatively small deviations below the threshold may lead to substantial bias for a specific regressor. Furthermore, the scores are likely to vary across different regressors as they capture the strength of the relationship between the cohort construction variables and the regressors.

Applying the threshold of 0.5 for the AWAR scores to all regressors leaves four cohort specifications which we present in Table 3.15. All four produce accurate estimates and it is clear the selection process has filtered out any specification which delivers heavily biased results. The selection process also retains the four best specifications from the full set of 65; Columns 2 and 3 contain the two best specifications where estimates for all regressions are within or close to the 95% confidence interval of the FE estimates. Columns 4 and 5 contain the next two best specifications with most estimates being within the 95% confidence intervals. Thus these cut-offs are stringent enough that they do not accept specifications which produce heavily biased estimates but nor are they too rigorous that they exclude specifications which deliver accurate estimates.

Shea's R-squared and the MES can be used as an additional check to the AWAR thresholds. Interestingly, using the recommended cut-off of 0.1 for the R-squared yields the same four specifications as a 0.5 cut-off for AWAR. All five specifications also have an MES of well over 100 and using a threshold closer to 130 would be consistent with the results from using the suggested cut-offs for AWAR and Shea's R-squared. The MES would be a good complement to use alongside AWAR when regressions include a large number of explanatory variables therefore imposing a threshold for each regressor may be overly cumbersome and instead should only be done for the key regressors of interest.

One potential concern with the model utilised is that variables related to income, such as secondary occupation and remittances, may be endogenous when estimating a consumption function. Thus the bias that is being addressed may not be due to aggregation but from endogeneity, which itself may be linked to cohort construction. To address this in Appendix B we re-estimate using a simplified consumption function where we drop the potentially endogenous secondary occupation and remittance variables. The results are almost identical for the simplified model, indicating aggregation is the main cause of the bias. One again sees that if there is insufficient cohort level variance pseudo-panel estimates are often worse than using simple OLS. Furthermore, using AWAR with a threshold of 0.5 again produces pseudo-panel estimates which are similar to fixed effects in a genuine panel.

Table 3.15: Pseudo-panel estimates passing the threshold tests

	Fixed Effects	Pseudo-Panel 4	Pseudo-Panel 5	Pseudo-Panel 6	Pseudo-Panel 7
HHsize	-0.090***	-0.087***	-0.089***	-0.090***	-0.091***
	(0.004)	(0.007)	(0.007)	(0.009)	(0.007)
ltot_asset	0.100***	0.112***	0.118***	0.095***	0.114***
	(0.006)	(0.013)	(0.012)	(0.015)	(0.011)
secondary	0.048***	0.057*	0.067**	0.058	0.088***
	(0.012)	(0.034)	(0.031)	(0.040)	(0.034)
lnremit	0.026***	0.044**	0.030*	0.065***	0.048***
	(0.007)	(0.018)	(0.015)	(0.020)	(0.016)
remittance	-0.283***	-0.517**	-0.332*	-0.747***	-0.515***
	(0.080)	(0.215)	(0.181)	(0.241)	(0.186)
Observations	9,510	9407	8583	9543	9324
Cohort Constructors		2 year, Region, Location, Education	2 year, Region, Location, Education, Gender	5 year, Region, Location, Education, Gender	2 year, Region, Education, Gender
Cohorts		392	492	316	405
Cell Size		6.5	5	8	6
AWAR					
HHsize		0.50	0.55	0.51	0.52
ltot_asset		0.60	0.63	0.58	0.60
secondary		0.51	0.60	0.50	0.53
lnremit		0.58	0.75	0.58	0.66
MES		148	169	131	154

Note: All regression include time dummies. The fixed effects regression trims the top and bottom 1 percentiles. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

7 Monte Carlo Simulations

The empirical application above has shown a lot of promise in using AWAR with a threshold of 0.5 as a test for the validity of cohort construction as well as its advantages over using IV-based measures. The fact that the threshold is consistent for all four regressors, which are very different in their structure and how they react to the presence of fixed effects indicates the results may be quite general. Nevertheless the empirical application is based only on one specific model using one specific dataset. More importantly, the data is panel in nature rather than repeated cross-sections which poses the potential for the results not being applicable to when genuine pseudo-panel estimation is required. In particular, as we have argued above, the panel nature may understate the measurement error problem and this will feed through into the AWAR measure as the ratio's denominator is essentially there to capture the potential extent of measurement error in the cohort sample means.

We conduct some Monte Carlo simulations in order to assess AWAR and its recommended thresholds to see how general the results from the empirical application are, particularly when using repeated cross-sections. The simulations indicate that using AWAR with a threshold of 0.5 works well in a number of situations, however the threshold is sensitive to cell sizes and the number of time periods. Therefore we develop another measure, called CAWAR, which adjusts for cell sizes and produce some tables containing recommended thresholds that vary with the number of time periods.

7.1 Testing AWAR and its Threshold

The Monte Carlo setup is similar to the ones used by Collado (1998) however we apply it to a linear model whereas they use it for a binary response mode. This setup is chosen because the data generating process used leads to a simple calculation and implementation of AWAR. It involves first generating the cohort population means (x_{ct}^{*}) as shown in equations 11 and 12, where recall:

$$x^*_{ct} = \rho x^*_{c(t-1)} + e_{ct} \qquad (11)$$

$$w_1 = \frac{\sigma_e^2}{1 - \rho^2} \qquad (12)$$

The initial period values of the cohort population means are generated as an IID $N(0, w_1)$ process where w_1 is as before the variance of the cohort population means. Similarly, e_{ct} is generated as IID $N(0, \sigma_e^2)$ and we discard the first ten cross-sections so we get w_1 values as shown in equation 12. Then the individual level explanatory variable can be generated as:

$$x_{it} = x^*_{ct} + v_{it} \qquad v_{it} \sim \text{i.i.d.} N(0, \sigma_v^2) \qquad (16)$$

By changing σ_e^2 and σ_v^2 we can change AWAR which is the square-root of the ratio of these two variances. For different values of AWAR we can then also vary ρ. We generate the unobserved heterogeneity which is correlated with the explanatory variable following Verbeek and Nijman (1992) as:

$$\theta_c = \lambda \bar{x}^*_c + \xi_c \qquad (17)$$

where $\lambda = 1, \bar{x}^*_c = \frac{1}{T}\sum_{t=1}^{T} x^*_{ct}, \xi_c \sim \text{i.i.d. } N(0,1)$ and $\text{E}\{\xi_c \left| x^*_{ct} \right.\} = 0$ for all $t = 1,..., \text{T}$

Finally we generate the dependent variable at the individual level as follows:

$$y_{it} = x^*_{ct}\beta + \theta_c + f_t + u_{it} \qquad (18)$$

where $\beta = 1$, f_t are time fixed effects and $u_{it} \sim \text{i.i.d.} N(0,1)$

We generate C cohort level observations, T number of time periods, and n_c number of individuals in each cohort. Thus in each time period we have $C \times n_c$ observations. All three dimensions will be altered during the course of the simulations, however as a baseline we will set T=4 (reflecting our empirical dataset) and n_c=50. We conduct 2000 repetitions for each set of simulations. Alongside with the AWAR values

which are calculated from the true parameter values for σ_e^2, σ_v^2 and ρ we also report the sample AWAR which uses the proxies from the individual level data as described in equation 14. In this way we can see if there are any significant differences between the true underlying AWAR values and the reported sample values. We vary the underlying AWAR values between 0.1 and 1 to identify suitable thresholds. We report the mean value of the coefficient estimates across the 2000 replications and the root-mean-squared-errors (RMSE). As the true coefficient value is set equal to unity, the mean estimates can be used to assess the average degree of bias (can interpret this in percentage terms by subtracting the mean estimate by unity) and the RMSE can be interpreted as showing the average absolute percentage error.

In Table 3.16 we show the results from varying the number of cohort level observations from 20 to 500. This range of cohort numbers is chosen as it likely captures the majority of possible pseudo-panel applications. The baseline setting of T=4 and n_c=50 is used, we set ρ=0.5 and alter AWAR by changing σ_e^2 while keeping σ_v^2=1. If we assume estimates are reasonably accurate when the mean values are within 10 percent of unity and RMSE is also less than 10 percent, then one can see a remarkable degree of consistency in the results suggesting that an AWAR value of 0.5 is the most suitable threshold for obtaining accurate estimates. Not only do we find evidence supporting the threshold level implied in the empirical application but also of the claim that the number of cohort level observations have little effect on the accuracy of pseudo-panel estimates, as was observed empirically in Figures 3.5a and 3.5b. The estimates perform slightly worse when there are only 20 cohorts but are very similar once there are at least 50. This suggests that it may be best to avoid using less than 20 cohorts but if it is unavoidable then a higher threshold may be required. The sample AWAR values are almost identical when the number of cohorts are varied, thus we just reported the average across all simulations in the second column. When AWAR is low the sample AWARs are over-estimated, however once it reaches 0.5 there is little difference between the true and sample values. In the final column we change the way AWAR is altered by changing σ_v^2 and keeping σ_e^2=1 instead while using 50 cohorts. Comparing it to the results with the

other set of simulations with 50 cohorts, one sees that only the relative value between σ_e^2 and σ_v^2, captured by AWAR, matters for estimation accuracy and not their individual magnitudes. We also find the 0.5 threshold holds irrespective of the degree of autocorrelation as shown in Table 3.17, although estimates are slightly more accurate when ρ is close to zero.

Once we begin to vary the number of time periods and the cell sizes (n_c), in Tables 3.18 and 3.19 respectively, the threshold value of 0.5 no longer appears to be a general result. Even when the number of time periods is varied it is only when T=2 or T=10 is a higher or lower threshold required. When T is between 3 and 8, as is the case for many of the repeated cross-sections survey datasets particularly for developing countries, a threshold of 0.5 is still optimal. In comparison, the required AWAR threshold is highly sensitive to the cell sizes, with each of the five cell sizes considered requiring a different threshold. Increasing cell sizes both lowers the true AWAR threshold required and reduces the upwards bias of the sample AWAR, further reducing the threshold one would need to use in empirical applications. One may question why in our empirical application a threshold of 0.5 was sufficient even though cell sizes were as low as 5 as Table 3.19 suggests that even with a cell sizes of 30 the threshold required needs to be at least 0.6. We believe this is due to our dataset being panel in nature which, as discussed above, would naturally reduce the measurement error problem as cohorts contain the same set of households over time. Thus cell size of 5 for a panel dataset could very well be equivalent to one of 50 for repeated cross-sections.

Table 3.16: Simulation Results for Varying the Number of Cohorts

AWAR	Sample AWAR	Number of Cohorts										Varying σ_v^2	
		20		50		100		200		500		50 Cohorts	
		Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
0.1	0.18	0.305	0.708	0.302	0.703	0.301	0.701	0.300	0.701	0.302	0.699	0.302	0.699
0.2	0.25	0.632	0.382	0.632	0.373	0.632	0.371	0.633	0.369	0.633	0.368	0.631	0.371
0.3	0.34	0.793	0.224	0.794	0.212	0.795	0.208	0.796	0.206	0.794	0.206	0.794	0.209
0.4	0.43	0.874	0.142	0.875	0.132	0.874	0.129	0.873	0.129	0.873	0.127	0.874	0.129
0.5	0.52	0.914	0.102	0.916	0.091	0.915	0.088	0.915	0.087	0.915	0.086	0.916	0.088
0.6	0.62	0.940	0.076	0.939	0.068	0.939	0.064	0.939	0.062	0.939	0.061	0.939	0.066
0.7	0.72	0.953	0.062	0.954	0.052	0.954	0.049	0.955	0.047	0.955	0.046	0.954	0.050
0.8	0.82	0.965	0.049	0.964	0.042	0.965	0.038	0.965	0.037	0.965	0.036	0.964	0.041
0.9	0.91	0.972	0.043	0.971	0.035	0.972	0.031	0.972	0.030	0.972	0.028	0.971	0.034
1	1.01	0.978	0.036	0.977	0.029	0.978	0.025	0.977	0.025	0.977	0.023	0.977	0.029

Note: The simulations set T=4, n_c=50, ρ=0.5. All simulations except those reported in the last column keep σ_v^2=1 and vary AWAR by changing σ_e^2. The last column does the reverse keeping σ_e^2=1 and changing σ_v^2 to vary AWAR. All simulations are conducted with 2000 replications. Sample AWAR are the average values across all simulations.

Table 3.17: Simulation Results for Varying the Degree of Autocorrelation

AWAR	Sample AWAR	Degree of autocorrelation (ρ)									
		0		0.2		0.5		0.7		0.9	
		Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
0.1	0.18	0.334	0.671	0.317	0.688	0.302	0.703	0.297	0.708	0.294	0.711
0.2	0.26	0.667	0.339	0.649	0.357	0.632	0.373	0.627	0.379	0.625	0.381
0.3	0.34	0.818	0.188	0.806	0.200	0.794	0.212	0.790	0.216	0.789	0.218
0.4	0.43	0.890	0.116	0.882	0.124	0.875	0.132	0.872	0.134	0.870	0.136
0.5	0.52	0.927	0.079	0.921	0.085	0.916	0.091	0.914	0.093	0.913	0.094
0.6	0.62	0.947	0.059	0.942	0.064	0.939	0.068	0.937	0.069	0.937	0.069
0.7	0.72	0.960	0.046	0.957	0.049	0.954	0.052	0.953	0.053	0.953	0.053
0.8	0.81	0.969	0.037	0.966	0.040	0.964	0.042	0.964	0.043	0.963	0.043
0.9	0.91	0.975	0.031	0.973	0.033	0.971	0.035	0.971	0.035	0.971	0.035
1	1.01	0.980	0.026	0.978	0.028	0.977	0.029	0.976	0.030	0.976	0.030

Note: The simulations set T=4, n_c=50, C=50, σ_v^2=1 and σ_e^2 is altered to vary AWAR. All simulations are conducted with 2000 replications. Sample AWAR are the average values across all simulations.

Table 3.18: Simulation Results for Varying the Number of Time Periods

AWAR	Sample AWAR	Number of Time Periods											
		2		3		4		5		8		10	
		Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
0.1	0.18	0.247	0.766	0.280	0.726	0.302	0.703	0.317	0.687	0.343	0.659	0.353	0.649
0.2	0.25	0.575	0.442	0.609	0.400	0.632	0.373	0.648	0.356	0.675	0.327	0.685	0.317
0.3	0.34	0.751	0.266	0.778	0.231	0.794	0.212	0.807	0.197	0.825	0.178	0.831	0.171
0.4	0.43	0.842	0.176	0.860	0.150	0.875	0.132	0.881	0.124	0.893	0.110	0.898	0.105
0.5	0.52	0.893	0.126	0.906	0.103	0.916	0.091	0.919	0.085	0.928	0.074	0.932	0.071
0.6	0.62	0.922	0.096	0.934	0.075	0.939	0.068	0.943	0.062	0.950	0.053	0.952	0.051
0.7	0.72	0.943	0.075	0.951	0.058	0.954	0.052	0.957	0.048	0.962	0.041	0.964	0.038
0.8	0.81	0.956	0.063	0.962	0.046	0.964	0.042	0.967	0.038	0.971	0.032	0.972	0.030
0.9	0.91	0.966	0.051	0.969	0.040	0.971	0.035	0.973	0.031	0.977	0.026	0.978	0.024
1	1.01	0.973	0.044	0.975	0.034	0.977	0.029	0.979	0.025	0.981	0.021	0.982	0.020

Note: The simulations set n_c=50, C=50, ρ=0.5, σ_v^2=1 and σ_e^2 is altered to vary AWAR. All simulations are conducted with 2000 replications. Sample AWAR are the average values across all simulations.

Table 3.19: Simulation Results for Varying the Cell Size

AWAR	Cell size														
	30			50			100			200			500		
	Sample AWAR	Mean	RMSE	Sample AWAR	Mean	RMSE	Sample AWAR	Mean	RMSE	Sample AWAR	Mean	RMSE	Sample AWAR	Mean	RMSE
0.1	0.21	0.206	0.798	0.18	0.302	0.703	0.15	0.462	0.543	0.13	0.634	0.372	0.11	0.810	0.196
0.2	0.28	0.508	0.497	0.25	0.632	0.373	0.23	0.773	0.233	0.21	0.872	0.134	0.20	0.944	0.062
0.3	0.36	0.699	0.306	0.34	0.794	0.212	0.32	0.886	0.120	0.31	0.940	0.066	0.30	0.974	0.031
0.4	0.45	0.804	0.202	0.43	0.875	0.132	0.41	0.932	0.074	0.41	0.964	0.042	0.40	0.986	0.020
0.5	0.54	0.866	0.141	0.53	0.916	0.091	0.51	0.956	0.051	0.50	0.977	0.029	0.50	0.991	0.014
0.6	0.64	0.903	0.104	0.62	0.939	0.068	0.61	0.969	0.037	0.60	0.984	0.022	0.60	0.994	0.012
0.7	0.73	0.928	0.079	0.72	0.954	0.052	0.71	0.977	0.029	0.70	0.988	0.017	0.70	0.995	0.009
0.8	0.83	0.944	0.063	0.82	0.964	0.042	0.81	0.982	0.024	0.80	0.991	0.014	0.80	0.997	0.008
0.9	0.93	0.954	0.053	0.91	0.971	0.035	0.90	0.986	0.020	0.90	0.993	0.012	0.90	0.997	0.007
1	1.03	0.963	0.044	1.01	0.977	0.029	1.00	0.989	0.017	1.00	0.994	0.011	1.00	0.998	0.006

Note: The simulations set T=4, C=50, ρ=0.5, σ_v^2=1 and σ_e^2 is altered to vary AWAR. All simulations are conducted with 2000 replications. Sample AWAR are the average values across all simulations.

7.2 Adjusting AWAR for Cell Size

The results have indicated that AWAR is a useful measure for determining the reliability of pseudo-panel estimates. However the threshold values required for making assessments varies across cell size, the number of time periods and, to a lesser degree, the level of autocorrelation. Thus for the measure to useful in applications one would need to produce a table of thresholds for each cell size, time period and potentially even ρ combinations. To avoid such a cumbersome endeavour we simplify the process by altering AWAR to incorporate cell sizes into the measure so that threshold values are stable over differing cell sizes. We call this measure cell size adjusted AWAR or CAWAR for short and its derivation is discussed below.

Recall that our AWAR measure is based on equation 13 where the denominator is the variance across individuals within a cohort. This is not the same as the measurement error variance, what Verbeek and Nijman call w_2, which was identified as the other factor alongside w_1 that in important in determining the accuracy of pseudo-panel estimates. To fully capture the problem of measurement error one must incorporate cell size alongside σ_v^2 as shown in equation 12. Combining equations 12 and 13 we can derive a ratio that fully accounts for measurement error, shown in equation 16.

$$w_2 = \sigma_v^2 \times n_c^{-1} \qquad (12)$$

$$\frac{\sigma_e^2}{\sigma_v^2} = \frac{w_1(1-\rho^2)}{\sigma_v^2} \qquad (13)$$

$$\frac{\sigma_e^2}{w_2} = \frac{w_1(1-\rho^2)n_c}{\sigma_v^2} \qquad (16)$$

Using the proxies identified before we can derive CAWAR as:

$$\text{CAWAR} = \frac{\hat{\sigma}_x^2(1-\hat{\rho}^2)n_c}{\hat{\sigma}_v^2} = (\text{AWAR})^2 \times n_c$$

In Tables 3.20a and 3.20b we show that our CAWAR measure with a threshold value of 12 can deliver accurate estimates (again defined as estimates with less than 10 percent bias both on average and in absolute terms) irrespective of cell size and cohort numbers. Particularly across cell sizes the estimation performance is almost identical whereas estimates are slightly poorer when there are less than 50 cohorts. The sample CAWAR values suggest that a slightly higher threshold of 13 may be required in actual applications using the proxies identified.

Table 3.20a: Varying Cell Size for CAWAR

Cell Size	CAWAR of 12		
	Sample CAWAR	Mean	RMSE
30	13.40	0.911	0.095
50	13.31	0.912	0.094
100	13.19	0.911	0.095
150	13.14	0.913	0.094
200	13.19	0.910	0.096
250	13.20	0.913	0.094
500	13.16	0.913	0.094
1000	13.13	0.913	0.094

Note: The simulations set T=4, C=50, ρ=0.5, σ_v^2=1 and σ_e^2 is altered to keep CAWAR=12 while n_c changes. All simulations are conducted with 2000 replications.

Table 3.20b: Varying the Number of Cohorts for CAWAR

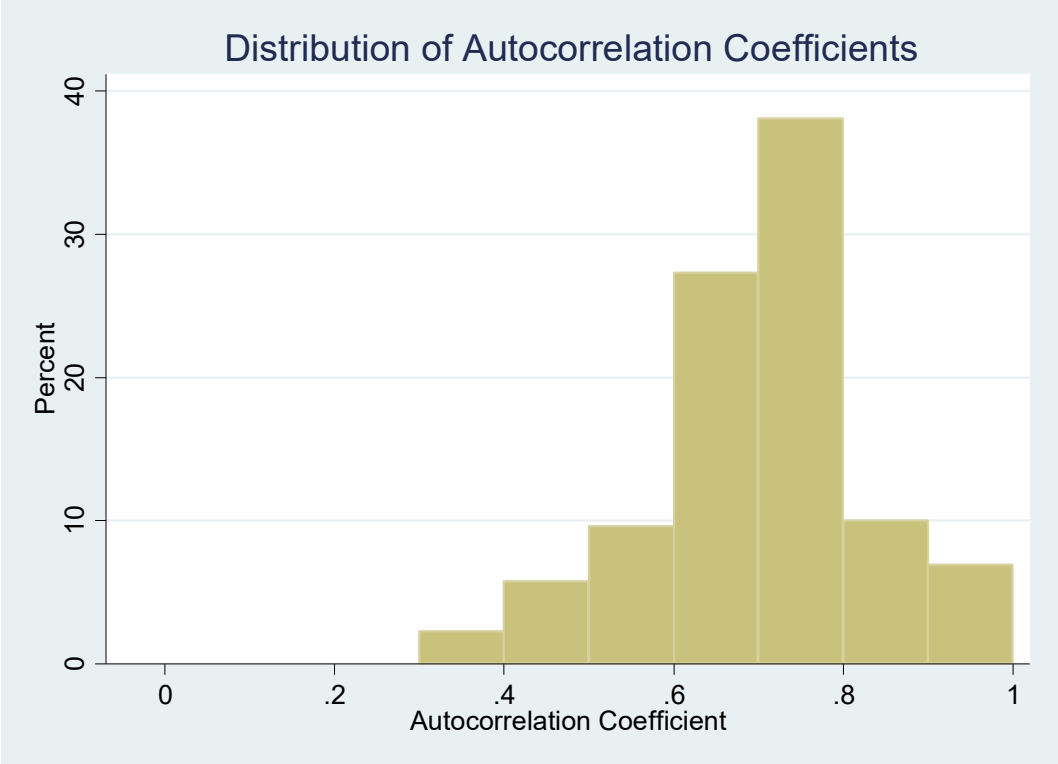
Number of Cohorts	CAWAR of 12		
	Sample CAWAR	Mean	RMSE
20	13.10	0.912	0.104
50	13.31	0.912	0.094
100	13.34	0.911	0.092
200	13.36	0.912	0.090
500	13.36	0.912	0.089

Note: The simulations set T=4, n_c=50, ρ=0.5, σ_v^2=1 and σ_e^2=0.24 CAWAR. All simulations are conducted with 2000 replications.

In Table 3.21 we present the CAWAR and sample CAWAR thresholds we recommend using for datasets with different time periods and autocorrelation coefficients. Figure 3.14 shows the distribution of autocorrelation coefficients for the

regressors of interest in our empirical application. One sees that most of the observations, about 75 percent, are distributed between the values of 0.6 and 0.9, although it should be noted that the coefficients are likely to be higher in our dataset as it is a panel than when repeated cross-sections are used. Nevertheless, using them as a guide for the potential autocorrelation coefficients one is likely to encounter, we see that when T is low then the CAWAR threshold required could be upward of 20 or more. Interestingly when T is larger, more than 5, a higher ρ reduces the required threshold whereas it increases for low T. When T is 4 or 5, the threshold is quite consistent over ρ. The thresholds are also consistent over T when the autocorrelation coefficient is lower than 0.4, especially when T≥4 the recommended threshold remains a 10 (11 using Sample CAWAR).

Figure 3.14: Distribution of Autocorrelation Coefficients from Empirical Application



We only consider up to 10 time periods as most repeat cross-sections are unlikely to extend beyond this amount, although the relevant thresholds can easily be found by expanding upon the simulations conducted. Given our applied dataset is panel in

nature we are unable to test the CAWAR thresholds empirically as the relationship between cell size and measurement error bias would be different given they are for repeated cross-sections. We thus leave this for future research. However the consistency of the AWAR threshold between the empirical application and Monte Carlo simulations should give cause for optimism.

Table 3.21: CAWAR Thresholds for Different Timer Periods

ρ	2 Time Periods				3 Time Periods			
	CAWAR	**Sample CAWAR**	**Mean**	**RMSE**	**CAWAR**	**Sample CAWAR**	**Mean**	**RMSE**
0	14	14.86	0.932	0.086	12	12.98	0.924	0.085
0.2	14	14.92	0.922	0.097	12	13.01	0.915	0.095
0.4	18	18.76	0.926	0.092	12	13.18	0.907	0.103
0.6	20	20.71	0.924	0.094	14	15.26	0.915	0.095
0.9	20	22.51	0.915	0.104	14	15.59	0.906	0.104

ρ	4 Time Periods				5 Time Periods			
	CAWAR	**Sample CAWAR**	**Mean**	**RMSE**	**CAWAR**	**Sample CAWAR**	**Mean**	**RMSE**
0	10	11.04	0.909	0.098	10	11.09	0.910	0.095
0.2	10	11.08	0.902	0.104	10	11.11	0.904	0.101
0.4	10	11.18	0.897	0.109	10	11.22	0.903	0.102
0.6	12	13.39	0.910	0.096	10	11.43	0.903	0.102
0.9	12	13.63	0.909	0.098	10	11.78	0.906	0.099

ρ	8 Time Periods				10 Time Periods			
	CAWAR	**Sample CAWAR**	**Mean**	**RMSE**	**CAWAR**	**Sample CAWAR**	**Mean**	**RMSE**
0	10	11.08	0.909	0.094	10	11.06	0.909	0.093
0.2	10	11.13	0.907	0.096	10	11.12	0.908	0.095
0.4	10	11.20	0.910	0.093	10	11.23	0.913	0.089
0.6	10	11.43	0.916	0.087	8	9.38	0.903	0.099
0.9	8	9.90	0.914	0.089	6	7.84	0.904	0.098

Note: The simulations set n_c=50, C=50, σ_v^2=1 and σ_e^2 is altered to vary CAWAR. All simulations are conducted with 2000 replications.

8 Alternative Estimation Techniques

In this section we consider alternative models and estimation techniques that incorporates cohorts but differs to the standard pseudo-panel estimation used thus far. We first see if bias arising from aggregating into cohorts can be avoided by keeping the data at the household level but utilising cohorts in some other way to account for the unobserved heterogeneity. We consider three different models but find none are suited to the task. We then look into some variations on pseudo-panel estimation which produce identical coefficient estimates but have different standard errors, commenting on how this effects inference.

We consider three different models which can incorporate cohorts into household level estimation: using cohort dummies in a household level OLS model, or using cohorts to allow the slope or intercept to vary across households. In the second column of Table 3.22 we present the results from estimating household level OLS including cohort dummies. Cohorts are constructed using 2 year age cohorts and all four of the construction variables as this combination gives the pseudo-panel estimates which are the closest to the "true" FE estimates as can be seen in Tables 3.4b and 3.15. Comparing the estimates to those in Column 3, which estimates using OLS as well but includes the cohort construction variables (age, region, location, gender and education) as controls, one sees little difference in the results with coefficient estimates being no closer to the "true" FE estimates in Column 1. Thus including cohort dummies is no better than just including the variables the cohorts are constructed from as additional regressors when using OLS.

An alternative way to incorporate cohort level information is to use mixed linear models which would allow for slope parameters to vary across households based on the cohort they belong to. It should be noted that if fixed effects are present then mixed linear models are inconsistent, however our purpose is to see if they can improve on OLS estimates when only repeated cross-sections are available. Column 4 estimates a random-intercept model which can be used to specify a different intercept for each household. If the intercept is specified to vary at the household

level then this would be identical to the RE estimator and as such would require panel data to estimate. Instead we allow the intercept to vary at the cohort level with the resultant estimates offering no improvements to estimating OLS with cohort dummies or cohort construction variables. The final column estimates a random-slopes model which allows the slopes of each of the regressors to vary by cohort. The estimates are very similar to the random-intercept model and there is again little benefit to using such a model over OLS. Thus from the three models considered there does not appear to be any suitable alternative to pseudo-panel estimation which can improve upon just using simple OLS (with construction variables as additional controls).

Table 3.22: Alternatives to Pseudo-Panel estimation

	FE	Pooled OLS	Pooled OLS	Random Intercept	Random Slope
HHsize	-0.090***	-0.094***	-0.096***	-0.098***	-0.104***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.003)
ltot_asset	0.100***	0.165***	0.169***	0.174***	0.176***
	(0.006)	(0.004)	(0.004)	(0.004)	(0.005)
secondary	0.048***	0.001	-0.010	-0.009	-0.014
	(0.012)	(0.011)	(0.010)	(0.011)	(0.012)
lnremit	0.026***	0.060***	0.067***	0.073***	0.078***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
remittance	-0.283***	-0.672***	-0.747***	-0.829***	-0.895***
	(0.080)	(0.077)	(0.073)	(0.072)	(0.077)
Other controls			Construction Variables		
Cohort Dummies		Yes			
Observations	9,510	9,510	9,510	9,510	9,510
R-squared	0.432	0.648	0.574		

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The focus so far has been on the bias of pseudo-panel estimates and ensuring they are as close as possible to the "true" fixed effects values. However, from Table 3.15 it is evident that the pseudo-panel estimates are also far less efficient than fixed effects with standard errors generally being at least two or three times larger. This has important consequences for inference as in a number of cases the pseudo-panel estimates are less significant (in one case losses significance completely) even

though the size of the coefficients are larger than for fixed effects. In some respect this may be inevitable due to the reduction in sample size when moving from the household to the cohort level. In Table 3.23 we consider some alternative ways of estimating pseudo-panels which produce the same coefficient estimates but different standard errors. The baseline method (which we have been using thus far) is shown in Column 3 where the data are aggregated to the cohort level and cohort dummies are included to account for fixed effects, what we term Cohort LSDV (least squares dummy variable). Cohorts are constructed using 2 year age bands and all four of the other construction variables as this specification yielded the closest estimates to the "true" FE estimates (which are included in Column 1 to aid comparison).

Table 3.23: Alternative techniques for estimating Pseudo-panels

Aggregation level	Household		Cohort		Household with cohort observations	
	FE	IV	Cohort LSDV	Cohort FE	LSDV Cohort Y	LSDV House-hold Y
HHsize	-0.090***	-0.089***	-0.089***	-0.090***	-0.089***	-0.089***
	(0.004)	(0.007)	(0.007)	(0.007)	(0.003)	(0.008)
ltot_asset	0.100***	0.118***	0.118***	0.118***	0.118***	0.118***
	(0.006)	(0.012)	(0.012)	(0.012)	(0.006)	(0.014)
secondary	0.048***	0.067**	0.067**	0.069**	0.067***	0.067*
	(0.012)	(0.032)	(0.031)	(0.032)	(0.016)	(0.037)
lnremit	0.026***	0.030*	0.030*	0.030**	0.030***	0.030
	(0.007)	(0.016)	(0.015)	(0.015)	(0.008)	(0.019)
remittance	-0.283***	-0.332*	-0.332*	-0.344*	-0.332***	-0.332
	(0.080)	(0.191)	(0.181)	(0.183)	(0.092)	(0.227)
Cohort Dummies	No	Yes	Yes	No	Yes	Yes
Observations	9,510	8,903	1,741	1,741	8,903	8,903
R-squared	0.432	0.656	0.908	0.685	0.914	0.555

Note: The first two columns estimate using household level data to either estimate fixed effects regression using the panel nature of the data or an IV regression using cohort means as instruments as described in equation 13. Columns 3 and 4 use cohort means to create a cohort level panel and account for cohort fixed effects either using cohort dummies (Column 3) or fixed effects (Column 4). Columns 5 and 6 maintains the data at the household level but the values of the explanatory variables are the cohort means. Column 5 uses the cohort means for the dependent variable whereas Column 6 instead uses the household level values. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The first method considered is the IV approach mentioned before and modelled in equation 13. As this approach keeps the data at the household level, one may expect the standard errors to be lower than for Cohort LSDV as there are more observations. This is not the case with standard errors for the IV being almost identical. For the second method we keep the data at the cohort level but use fixed effects estimation (labelled Cohort FE) rather than cohort dummies as this increases the degrees of freedom so potentially may reduce the size of the standard errors. The results in Column 4 indicate that if anything this makes the standard errors slightly larger. The coefficients are also slightly different as fixed effects estimates requires cell sizes (which are used as regression weights) to be constant over time for each cohort whereas LSDV allows them to vary. We also consider alternative techniques for estimating standard errors (not reported) using Cohort LSDV and FE. We find that clustered, Jack-Knife and Bootstrap standard errors are all similar or slightly larger. Bootstrapping could only be used for FE and does not allow for regression weights to be included so the coefficients estimates differ. Thus neither of the two methods nor the alternative techniques for calculating standard errors produce errors similar to the FE estimates.

The final two methods are similar and involve estimating at the household level but using the cohort mean observations rather than the household level observations. This amounts to the cohort mean observations being repeated in the dataset the same number of times as their cell sizes. Cohort dummies are included to account for fixed effects and the regression is weighted by the negative of the square-root of the cell size to make estimation consistent with the cohort level regressions. The methods differ in that "LSDV Cohort Y" uses cohort mean values for the dependent variable as the well the independent variables whereas "LSDV Household Y" does so only for the independent variables and keeps the dependent variable values at the household level. Both approaches have a certain advantage over using Cohort LSDV. Cohort Y produces standard errors which are very similar to the FE errors and this results in identical inference with estimates having the same significance levels. Although estimating in such a manner may seem questionable as it essentially involves artificially expanding the number of observations by repetition, it is the only method that delivers errors close to the FE values with all other methods failing

to improve on the Cohort LSDV errors. Thus it would be best to continue using Cohort LSDV as the baseline as its estimation is more justifiable, however one could include Cohort Y alongside it in order to get inference closer to those at the household or individual level. The Household Y approach is not as useful for getting improved standard errors, but has the advantage of producing an R-squared which is more consistent with those from household level models. Although the R-squared is still larger than for the FE model, it is far below those from other pseudo-panel models where the value often exceeds 0.9. Also it is similar to, in fact smaller than, the R-squared for household level OLS estimates in Table 3.2. Thus if one is interested in a goodness-of-fit measure for pseudo-panel models that is more consistent with household or individual level data then the R-squared from a Household Y model is a suitable option. The estimation process itself is also justifiable as it does not run into the problem of artificially expanding observations by repetition as the dependent variable remains at the household level.

9 Conclusion

The main concern regarding pseudo-panel estimation of linear fixed effects models has traditionally been the bias arising from measurement error caused by the cohort sample means being poor estimates of the underlying cohort population means. Estimating pseudo-panel regressions using genuine panel data minimises such bias, nevertheless coefficient estimates differ greatly based on how cohorts are constructed; with some being almost identical to panel fixed effects estimates while others are more biased than OLS. We find that, contrary to expectations, pseudo-panel estimates are less biased when cell sizes are lower and consequently the number of cohort level observations are higher. We explain this result as a consequence of bias arising from the process of aggregating household into cohorts which can lead to the loss of importance sources of variation and heterogeneity in the data. Furthermore, we find some of the methods commonly used for constructing cohorts in the literature are unsuitable as they do not account for this important

source of bias, particularly when the level of analysis is at the household or individual level.

To address this, we focus our attention on the cohort construction process in order to provide some practical guidance on how it should be conducted. We find that age is the most important variable to include in any cohort specification, in particular age bands of 5 years or less. For the other construction variables it is more difficult to judge which ones are better to include as it is their performance in combination with each other that drive the results. Thus we shift our attention to constructing a measure for assessing how likely it is that the overall cohort specification will deliver consistent pseudo-panel estimates. The measure assesses the degree of three types of variation in the cohort level data which have been shown to be important in the literature: across variation which captures the variability in the cohort level data, within variation which captures the variability of individuals that have been grouped into each cohort, and time variation which captures the degree of serial correlation in the cohort data. We identify some suitable thresholds for accepting or rejecting cohort specifications based on this measure, which we call AWAR, both using an empirical application and Monte Carlo simulations. The simulations show that thresholds vary greatly across cell sizes and the number of time periods. Therefore we introduce an adjustment for cell size, calling the adjusted measure CAWAR, which keeps the recommended threshold constant over cell size but still varies across the number of time periods and the degree of autocorrelation. We produce some tables containing recommended CAWAR thresholds for different time period and autocorrelation combinations which are relevant for empirical applications. We also highlight the similarity of cohort construction to instrument selection and demonstrate how some of the tests for instrument validity can be applied to the pseudo-panel context.

We believe that CAWAR and its recommended thresholds should be the starting point of verifying results from pseudo-panel estimation. Further work is required in order to confirm the suitability of these thresholds, particularly from empirical applications as we have only tested CAWAR using Monte Carlo simulations. The consistency of the AWAR thresholds in both the empirical application and Monte

Carlo simulations do provide evidence of empirical validity but they are limited by the fact it is based on estimates from panel data. One way to empirically validate the thresholds is to use a large panel data set where the "true" coefficient can be estimated using panel fixed effects. Then a random subset of the population can be drawn for each time period to create a repeated cross-section with which pseudo-panels can be estimated. The dataset would need to be large enough that the subsets in each time period have enough observations to produce CAWAR that can meet the thresholds, which based on our empirical application may mean each subset requires at least a thousand observation.

The thresholds may also be sensitive to the model being estimated and further simulations are required to address this. Pseudo-panel estimation has been expanded beyond the linear fixed effects model considered here and applications include dynamic and nonlinear models. Such models may require greater level of variation at the cohort level for consistency or be more sensitive to measurement error than a simple linear model thus would need a higher CAWAR threshold. The next chapter considers one such nonlinear model which contains a binary response variable. Another concern would be for models which incorporate other types of heterogeneity, such as one where the parameters vary across households. In this case greater variation is likely to be needed at the cohort level in order to capture the full range of households across which parameter values differ. Consequently, for application of pseudo-panel estimation to new and more complex models, one may need to first use simulations to calibrate the required threshold before conducting the empirical analysis.

Appendix to Chapter 2

Appendix A

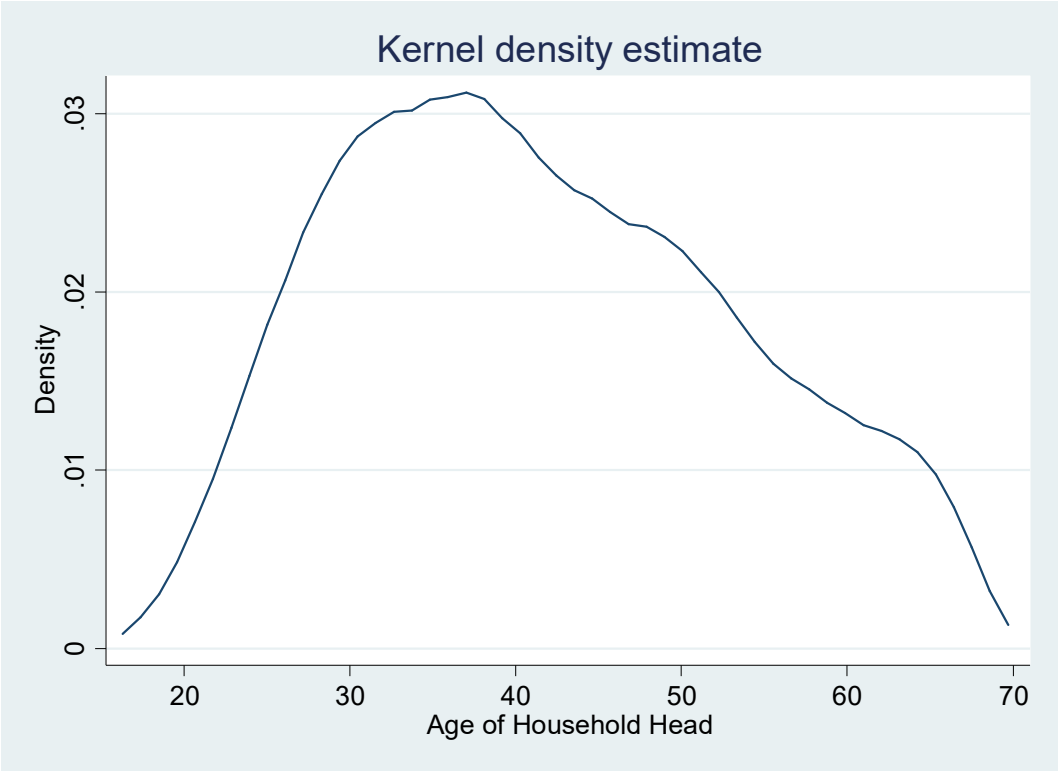
Table 3.A1: Summary Statistics of Main Variables

	Log Consumption			**Log Total Assets**			**Log Remittances**		
	2005/06	2011/12	Total	2005/06	2011/12	Total	2005/06	2011/12	Total
Mean	10.70	11.41	11.08	13.40	14.31	13.97	11.14	12.57	11.91
Variance	0.62	0.58	0.67	3.11	3.46	3.64	3.32	1.97	3.06
Median	10.62	11.36	11.03	13.26	14.23	13.90	11.08	12.47	11.92
Obs	2851	2488	9966	2851	2488	9966	1086	646	2938
	Household Size			**Secondary Occupation**			**Received Remittance**		
	2005/06	2011/12	Total	2005/06	2011/12	Total	2005/06	2011/12	Total
Mean	5.22	5.47	5.52	0.39	0.47	0.45	0.38	0.26	0.29
Variance	8.50	7.32	8.04	0.24	0.25	0.25	0.24	0.19	0.21
Median	5	5	5	0	0	0	0	0	0
Min	1	1	1	0	0	0	0	0	0
Max	29	21	29	1	1	1	1	1	1
Obs	2851	2488	9966	2851	2488	9966	2851	2488	9966

Table 3.A2: Summary Statistics for Age of Household Head

	Age				
	2005/06	**2009/10**	**2010/11**	**2011/12**	**Total**
Mean	39.00	42.68	42.64	42.68	41.61
Variance	147.99	131.75	133.86	136.38	140.75
Median	37	41	42	41	40
Obs	2851	2334	2293	2488	9966

Figure 3.A1: Distribution of the Age of the Household Head



Appendix B

In the empirical application one may be concerned that the variables related to income, secondary occupation and the remittance dummies, may be driving the results not due bias from aggregation but from endogeneity as we are using a consumption function. Thus if the endogeneity is also linked to the cohort construction process, what we have been addressing in the analysis is not bias arising from aggregation but from endogeneity. In order to alleviate such concerns we perform some robustness checks using a simplified model that excludes the potentially endogenous secondary occupation and remittances variables. As such we can be confident that bias coming from the cohort construction process is due to aggregation and loss of cohort level variation rather than endogeneity.

We begin by reproducing Tables 3.4a and 3.4b using the simplified model that excludes secondary occupation and the two remittance variables, the results are presented in Tables 3.B1 and 3.B2 respectively. Table 3.B1 shows that even with the simplified model many of the common cohort specifications (particularly those advocating just using specifications that result in cell sizes 100-200) produce pseudo-panel estimates which are often worse than Pooled OLS. However Table 3.B2 shows that, as before, if the correct cohort specification is used pseudo-panels can deliver estimates just as good as fixed effects from a genuine panel. Consequently one sees that cohort construction does still matter greatly even when potentially endogenous variables are removed.

Figure 3.B1 reproduces Figure 3.9 which shows the relationship between estimation bias and the AWAR value for the whole set of 65 cohort specifications but using the simplified model. Again a strong negative relationship is clear and the AWAR threshold of 0.5 above which estimates have much less bias is apparent for both the household size and log total assets variables. Figure 3.B1 also illustrates that the pattern of aggregation bias highlighted in the main analysis is retained for the simplified model across the 65 cohort specifications and not just the handful shown in Tables 3.B1 and 3.B2. Finally, Table 3.B3 shows the resulting pseudo-panel estimates when we use the cohort specifications which meet the 0.5 threshold for all

variables, the simplified model version of Table 3.15. All four specifications which meet the threshold produce pseudo-panel estimates that are almost identical to fixed effects. Thus our prescribed test for assessing cohort construction works just as well for the simplified model, meaning that it is addressing bias from aggregation not endogeneity.

Table 3.B1: Estimating Table 3.4a using the Simplified Model

	Pooled OLS	Fixed Effects	Pseudo-Panel 1	Pseudo-Panel 2	Pseudo-Panel 3	Pseudo-Panel 4
HHsize	-0.093***	-0.090***	-0.068***	-0.077***	-0.025	-0.022
	(0.003)	(0.004)	(0.015)	(0.015)	(0.028)	(0.085)
ltot_asset	0.174***	0.102***	0.166***	0.185***	0.083	0.095
	(0.005)	(0.006)	(0.035)	(0.025)	(0.069)	(0.085)
Other Controls	Yes	No	No	No	No	No
Observations	9530	9,533	9,855	9,837	9,939	9,966
Cohort Constructors			2 year	2 year, Region	5 year, Location	Region, Education
Cohorts			25	96	20	12
Cell Size			100	26	124	208

Note: As Table 3.4a

Table 3.B2: Estimating Table 3.4b using the Simplified Model

	Pooled OLS	Fixed Effects	Pseudo-Panel 4	Pseudo-Panel 5	Pseudo-Panel 6
HHsize	-0.093***	-0.090***	-0.088***	-0.088***	-0.088***
	(0.003)	(0.004)	(0.007)	(0.007)	(0.009)
ltot_asset	0.174***	0.102***	0.115***	0.120***	0.098***
	(0.005)	(0.006)	(0.013)	(0.012)	(0.016)
Other Controls	Yes	No	No	No	No
Observations	9530	9,533	9407	8583	9543
Cohort Constructors			2 year, Region, Location, Education	2 year, Region, Location, Education, Gender	5 year, Region, Location, Education, Gender
Cohorts			363	435	291
Cell Size			6.5	5	8

Note: As Table 3.4b

Figure 3.B1: Reproduction of Figure 3.9 using the Simplified Model

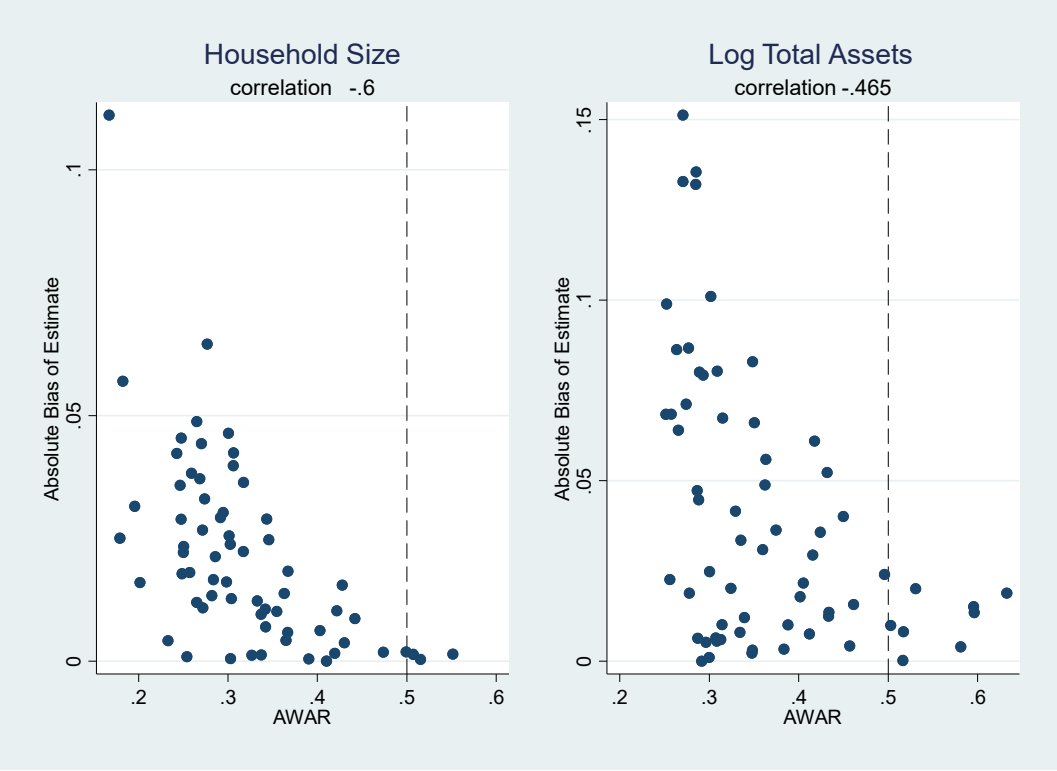


Table 3.B3: Estimating Table 3.15 using the Simplified Model

	Fixed Effects	Pseudo-Panel 4	Pseudo-Panel 5	Pseudo-Panel 6	Pseudo-Panel 7
HHsize	-0.090***	-0.088***	-0.088***	-0.088***	-0.090***
	(0.004)	(0.007)	(0.007)	(0.009)	(0.007)
ltot_asset	0.102***	0.115***	0.120***	0.098***	0.117***
	(0.006)	(0.013)	(0.012)	(0.016)	(0.012)
Observations	9,533	9407	8583	9543	9324
Cohort Constructors		2 year, Region, Location, Education	2 year, Region, Location, Education, Gender	5 year, Region, Location, Education, Gender	2 year, Region, Education, Gender
Cohorts		392	492	316	405
Cell Size		6.5	5	8	6
AWAR					
HHsize		0.50	0.55	0.51	0.52
ltot_asset		0.60	0.63	0.58	0.60
MES		384	435	354	387

Note: As for Table 3.15

Chapter 4:

Pseudo-Panel Estimation of Binary Response Models

Abstract

This chapter assesses four different estimation methods of binary response models with individual effects where the data is a time-series of independent cross-sections. We compare Deaton's (1985) linear fixed effects estimator, which is most often used in applied work, to three non-linear estimators. The first is a simplified version of Collado's (1998) Minimum Distance estimator. The other two are based on the fractional response estimators developed by Wooldridge and Papke (2008) which, unlike Collado's estimator, can be used for dynamic models as well. Results from Monte Carlo simulations and from an empirical application indicate the linear estimator is just as good as the other non-linear estimators and is generally more robust to problems arising from the process of creating the pseudo-panel.

1 Introduction

The previous chapter highlighted the value of developing suitable pseudo-panel estimation techniques for analysis in developing countries where panel data are unavailable, particularly for medium to long run analysis. An important extension to this is the inclusion of binary response models given that many of the key concerns in development economics are of this nature. This includes but is not limited to; poverty analysis, employment status, health outcomes, educational attainment and consumption or access to certain resources/services. In most of these cases the presence of unobserved heterogeneity at the household or individual level will be crucial, requiring the use of fixed effects and thus pseudo-panel estimation, given the high likelihood that repeated cross-sections rather than panel data are all that is available. However, developments in the literature on estimating binary response pseudo-panel models have been limited thus far.

Moffitt (1993) proposes a general way of estimating with repeated cross-sections based on an instrumental variables approach where cohort dummies interacted with time are used as instruments as an alternative to the cohort means approach of Deaton (1985). Such an approach can be extended to binary response models: Collado (1998) uses Deaton's method and combines it with a correlated random effects approach based on Chamberlain (1980, 1982) for binary response panel data models. However neither of these methods have been used in empirical applications, where the general approach has been to continue using the linear fixed effects pseudo-panel estimator, i.e. the EWALD estimator discussed in previous chapters, and ignore the discrete nature of the dependent variable. Gassner (1998) and Bedi et. al. (2004) use the linear estimator for their models on telephone access in the UK and school enrolment in Kenya. Dargay and Vythoulkas (1999) and Jæger (2013) also estimate in this way, although their dependent variable is a count rather than binary.

This raises the question of how well the linear estimator performs when applied to nonlinear pseudo-panel models as it is so commonly used in empirical applications. In panel data models the linear estimator, known as the Linear Probability Model

(henceforth LPM) which uses OLS with fixed effects, is often found to be adept at estimating average marginal effects (henceforth AME). It also has the advantage of leaving the correlation between the unobserved heterogeneity and the explanatory variables (the essential cause for the need for including fixed effects) unspecified, whereas nonlinear estimators may need to assume a specific correlation. Whether these properties also hold for pseudo-panel estimation requires investigation due to the additional concerns regarding cohort level variation and measurement error bias discussed in Chapter 3. In particular, bias caused by a lack of cohort level variation may be exacerbated due to reduced variation in the dependent variable as it is now binary rather than continuous. Additionally, a nonlinear estimator suited to binary response models may suffer less from this problem. It is also unclear for such models whether a linear or nonlinear estimator would suffer more from measurement error.

Consequently this chapter has two main motivations. Firstly, to assess the adequacy of the linear fixed effects estimator when using binary pseudo-panel models. This not only allows us to verify the results from previous empirical studies which used this method but also validate future applications where the linear estimator may be favourable to nonlinear ones. Reasons for this favourability could be due to the linear estimator being able to estimate marginal effects without assuming a specific correlation about the fixed effects, or wanting to use some of the extensions to the linear pseudo-panel estimator such as those which correct for measurement error (see Deaton, 1985; Verbeek and Nijman, 1993; Devereux 2007a), which are useful when cell sizes are low. Secondly, to compare linear and nonlinear pseudo-panel estimators and identify scenarios where one is likely to outperform the other.

For this second purpose it is necessary to augment and expand on the existing nonlinear estimator suggested by Collado. The method essentially involves grouping the explanatory variables into cohorts and calculating their means but leaving the dependent variable at the individual level. Then for each time period one regresses the binary dependent variable on all leads and lags of the cohort means of the explanatory variables using Probit estimation. Finally, minimum distance is used on the set of parameter estimates to calculate the structural parameters. We use a simplified version based on Mundlak (1978) where the time averages of the explanatory variables rather

than all their leads and lags are included as additional regressors. This approach conserves on degrees of freedom, does not require the additional step of minimum distance estimation and allows for the average marginal effects, which are of more interest for empirical interpretation as opposed to coefficient estimates, to be identified very generally. As a result the Mundlak version has been more widely used in the panel data literature, with extensions developed for Tobit, count, and fractional response models.

This latter set of extensions are useful for our purposes as when a binary dependent variable is aggregated to the cohort level using cohort means it becomes fractional. One problem with Collado's method, and our simplified version, is that it leaves the dependent variable at the individual level and uses cohort level data for the explanatory variables. Apart from this being a less than intuitive approach to estimation it also precludes the possibility of extending the model using specifications that require the dependent variable to be traceable over time. For example models which include dynamic specifications, such as using lagged dependent variables, or are estimated in differences cannot be used with this estimation procedure. However, using the two fractional response model estimators developed by Papke and Wooldridge (2008) we can aggregate the dependent variable to the cohort level, thus making it traceable over time, and still proceed with nonlinear estimation which takes into consideration the boundedness of the dependent variable. Thus we have three nonlinear estimators which can be tested against the linear estimator; one based on Collado's approach and two on fractional response models.

We test the small sample properties of these four estimators using Monte Carlo simulation and compare the effect of changing cohort level variation, measurement error, and autocorrelation in the explanatory variable. The simulations are designed in a similar fashion to those conducted by Collado with the exception that we report the average marginal effects rather than the coefficient estimates. This not only allows for comparison across linear and nonlinear models but is also of more interest for economic analysis. We then use four waves of data from the Uganda National Panel Survey to check how the pseudo-panel estimators perform compared to genuine panel estimates in a similar fashion to Chapter 3.

Section 2 gives an overview of the literature on binary response panel data models, focusing on the problems associated with the various different ways of accounting for fixed effects using both linear and nonlinear models. Section 3 explains the derivation of the four pseudo-panel estimators to be used. Section 4 details the Monte Carlo simulation procedure and discusses the results. Section 5 covers the empirical application using the Ugandan data; discussing the data, methodology, and results. In section 6 we discuss how the estimators can be extended to allow for endogenous regressors and whether lagged variables from the cohort level observations can be used as valid instruments. Finally, section 7 concludes.

2 Binary Response Panel Data Models

In this section we will be discussing the estimation of a static fixed effects model where the dependent variable is binary. In the panel literature, estimating nonlinear fixed effects models has been found to be rather problematic. In the linear case one can get rid of fixed effects either by differencing, taking a within transformation (demeaning the variables) or including dummy variables for each individual. However none of these methods are in general applicable to nonlinear estimators. To see why, consider the probit model in equation (1) with fixed effects represented by f_i, Φ is the standard normal cumulative distribution function and $\boldsymbol{x}_{it}$ are the explanatory variables. Neither differencing nor demeaning the variables will eliminate the fixed effects given the structural form of the model. Next, consider the marginal effects of such a model, equation (2) derives the marginal effect for the j^{th} explanatory variable where ϕ is the standard normal density function. The marginal effects depend not only on the explanatory variables and their coefficients but also on the fixed effects. Thus to get accurate estimates of the marginal effects and coefficients, accurate estimates of the fixed effects are required. Using dummies for each individual to account for the fixed effects means the latter are replaced by sample estimates, which are inconsistent unless the number of time periods is large, causing bias in the coefficient and marginal effect estimates. This is the well-known incidental parameters problem identified by

Neyman and Scott (1948). Although there have been some attempts at obtaining bias corrected versions of this dummy variable approach these often require larger number of time periods than is usually available, especially for household surveys, see Hahn and Newey (2004) and Fernández-Val (2009).

$$\mathrm{P}(y_{it}=1 \mid \boldsymbol{x_{it}}, f_i) = \Phi(\boldsymbol{x_{it}}\boldsymbol{\beta} + f_i), \quad i = 1,...,N \quad t = 1,...,T \qquad (1)$$

$$\frac{\delta \mathrm{E}(y_{it} \mid \boldsymbol{x_{it}}, f_i)}{\delta \boldsymbol{x_{itj}}} = \beta_j \phi(\boldsymbol{x_{it}}\boldsymbol{\beta} + f_i) \qquad (2)$$

Alternatively one could just use the linear OLS estimator, also known as the LPM, which can easily account for the fixed effects. Doing so assumes a model of the form given in equation (3), where the probability of success is assumed to be a linear function of the independent variables and the fixed effects. The marginal effects of this model are comparatively easier to estimate and interpret as shown in equation (4). The major problem with this model is that the predicted probabilities may not lie between zero and one, making such predictions meaningless for interpretation. Similarly, from equation (4) one can see that the marginal effects are assumed constant irrespective of the value of the explanatory variables. This is problematic as it implies continually increasing the explanatory variables will drive the probability predictions below zero or above one. Nevertheless, the LPM can generally be viewed as a reasonable linear approximation which often delivers reliable estimates of average marginal effects.

$$\mathrm{P}(y_{it}=1 \mid \boldsymbol{x_{it}}, f_i) = \boldsymbol{x_{it}}\boldsymbol{\beta} + f_i, \quad i = 1,...,N \quad t = 1,...,T \qquad (3)$$

$$\frac{\delta \mathrm{E}(y_{it} \mid \boldsymbol{x_{it}}, f_i)}{\delta \boldsymbol{x_{itj}}} = \beta_j \qquad (4)$$

There are a few nonlinear panel estimators which account for the unobserved heterogeneity in different ways. One is the fixed effects logit estimator or the conditional logit estimator. Assuming a logit functional form, one can find a conditional distribution that does not depend on f_i and hence get consistent estimates

of the coefficient using conditional maximum likelihood estimation (CMLE). This method has the advantage that it can leave the correlation between the unobserved heterogeneity and the explanatory variables unspecified but it cannot be applied to probit models. Furthermore, marginal effects and AMEs cannot in general be estimated using this method as it requires either plugging in values for f_i or specifying a distribution for it. This means one cannot estimate the absolute magnitudes of the effects of each covariate but it is possible to see the direction of the effect and estimate relative magnitudes of covariates by, for example, calculating the log-odds ratio.

The inability to estimate marginal effects will be a major drawback in most empirical applications. Thus it is necessary to continue searching for estimators which can. Random effects estimators are relatively simple to estimate for nonlinear models and marginal effects can easily be obtained. However they impose a severe restriction by assuming that the unobserved heterogeneity is independent of the explanatory variables, shown in equation (5), resulting in biased estimates when fixed effects are present. Chamberlain (1980) allowed for dependence between f_i and the covariates by assuming a specific correlation between them. Mundlak's (1978) version of this assumption is shown in equation (6) where $\tilde{\boldsymbol{x}}_i$ is the time average of $\boldsymbol{x}_{it}$ and $t = 1,\ldots,T$. This amounts to assuming the fixed effects are a linear function of the time means of the covariates, normally distributed and with a constant variance. Chamberlain's version allows a bit more generality, shown in equation (7), by replacing the time means with all the time observation for the covariates. Essentially, it allows the λ's in (7) to differ whereas the Mundlak version assumes they are the same across all time periods.

$$f_i \mid \boldsymbol{x}_{it} \sim \text{Normal}(0,\sigma_c^2) \tag{5}$$

$$f_i \mid \boldsymbol{x}_{it} \sim \text{Normal}(\psi + \tilde{\boldsymbol{x}}_i\xi,\sigma_a^2) \tag{6}$$

$$f_i \mid \boldsymbol{x}_{it} \sim \text{Normal}(\psi + \boldsymbol{x}_{i1}\boldsymbol{\lambda}_1 + \ldots + \boldsymbol{x}_{iT}\boldsymbol{\lambda}_T,\sigma_a^2) \tag{7}$$

The estimators based on the assumptions in equations (6) and (7) are known as the correlated random effects (CRE) estimators. They are usually combined with a probit specification, although logit is also possible, and marginal effects can easily be derived. Most applications use the simpler Mundlak version, which conserves on

parameters and can lead to a simple probit model shown in equation (8), which combines the model in equation (1) with assumption shown in equation (6), where the a subscripts denote the parameters have been multiplied by $(1+\sigma_a^2)^{-1/2}$, see Wooldridge (2010a) for more details. The Chamberlain version requires the model to be estimated separately for each time period where all leads and lags of all covariates are included. Then minimum distance is used to derive the structural parameters using the reduced form estimates of each time period regression.

$$\mathrm{P}(y_{it}=1 \mid \boldsymbol{x}_{it}) = \Phi(\psi_a + \boldsymbol{x}_{it}\boldsymbol{\beta}_a + \tilde{\boldsymbol{x}}_i\xi_a), \quad i=1,...,N \quad t=1,...,T \tag{8}$$

Consequently CRE, specifically the Mundlak version, is often the most popular estimator to use for binary response panel data to account for fixed effects. It has the disadvantage that a specific correlation between the unobserved heterogeneity and the covariates needs to be assumed, whereas the LPM and the fixed effects logit estimator require no such assumption. However, as the latter estimator cannot generally be used to estimate marginal effects its use may be limited to that of a robustness check on the sign and relative magnitudes of other estimators. The LPM suffers from the fact it cannot account for the binary nature of the response variable and hence it may produce probability predictions that are not meaningful for interpretation, although in many cases it can provide useful approximations of the overall marginal effect. Therefore it can be useful to use both LPM and CRE as neither is unambiguously better than the other. One exception is if one is interested in marginal effects at specific points or at specific values of a covariate, these can only be estimated with nonlinear estimators as the LPM just assumes a constant marginal effect. With CRE marginal effects at specific values can be estimated but under the strong assumption of conditional independence in the response variable. In contrast AMEs are simpler to estimate as pooled probit can be used, which drops completely the conditional independence assumption.

The Mundlak version of the CRE approach has been extended to other nonlinear fixed effects models including Tobit, count, and fractional response models. There have also been extensions which allow for it to be estimated under unbalanced panels

(Wooldridge 2010b) and to include instrumental variables (Papke and Wooldridge 2008). Although many of these extensions can also be applied to the Chamberlain version, it is generally Mundlak's which is favoured. Thus we also use the Mundlak version for our binary response pseudo-panel estimator in contrast to Collado. Furthermore, we develop two additional pseudo-panel estimators based on the fractional response model developed by Papke and Wooldridge (2008) using the Mundlak assumption. Thus the three nonlinear estimators we propose have a consistent method for dealing with fixed effects, allowing us to focus on differences arising from the estimation technique rather than the assumptions underlying the unobserved heterogeneity.

3 Pseudo-Panel Estimators of Binary Response Models

We now detail the estimation process of the linear and three nonlinear pseudo-panel estimators we propose using to estimate binary response models when the data are repeated cross-sections. Firstly, the linear estimator which is a simple application of Deaton's (1985) method which was used in the previous chapter, known as the EWALD estimator. This involves grouping individuals into cohorts, based on time invariant characteristics such as year of birth, and using the cohort means as if they were observations in a real panel. Then cohort dummies are included to account for fixed effects. The model is show in equation (9), where $\bar{y}_{ct}$ is the cohort mean of the binary dependent variable which becomes a proportion at the cohort level, $\bar{\boldsymbol{x}}_{ct}$ is the cohort means of the explanatory variables, and f_c is the cohort fixed effects which are correlated with the explanatory variables and are accounted for using cohort dummies.

$$\bar{y}_{ct} = \bar{\boldsymbol{x}}_{ct}\boldsymbol{\beta} + f_c + \varepsilon_{ct}, \quad c = 1,...,C \quad t = 1,...,T \qquad (9)$$

The second estimator, which henceforth we will refer to as CRE, is based on Collado's estimation process but using the Mundlak version rather than Chamberlain's. This involves regressing the cohort means of the explanatory variables as well as their averages across time on the individual level dependent variable. The model to be

estimated using probit is shown in equation (10), note here we assume the constant term is absorbed into $\bar{\boldsymbol{x}}_{ct}$. The individual binary dependent variable is represented by y_{it}, $\bar{\boldsymbol{x}}_{ct}$ is as before and $\tilde{\bar{\boldsymbol{x}}}_c$ are the cohort level time averages of the explanatory variables.

$$\mathrm{P}(y_{it}=1 \mid \bar{\boldsymbol{x}}_{ct}) = \Phi(\bar{\boldsymbol{x}}_{ct}\boldsymbol{\beta} + \tilde{\bar{\boldsymbol{x}}}_c\xi), \quad i=1,...,N \quad c=1,...,C \quad t=1,...,T \qquad (10)$$

Alternatively, one may find it more intuitive or necessary to estimate a model where all the variables are at the same level of aggregation. The final two estimators allow for this while still using nonlinear estimation. The estimators use a model similar to that in equation (10), however, as at the cohort level the dependent variable is now fractional rather than binary a straightforward application of CRE is not possible. Instead what can be used is the estimation developed by Papke and Wooldridge (2008) who apply the CRE approach, using the Mundlak version, to a fractional response model. They develop two estimators which can be used to calculate coefficients and marginal effects. The first estimates using the pooled Bernoulli quasi-MLE estimator, which can be obtained by maximising the pooled probit log-likelihood, which they refer to as the "pooled fractional probit" (or PFP) estimator. The second uses a panel generalised estimating equation (GEE) to estimate a probit model with an exchangeable within-group correlation structure. The authors also recommend using fully robust bootstrap standard errors for the marginal effects of both PFP and GEE.

It is worth noting that the AMEs for all three nonlinear estimators discussed above; CRE, PFP and GEE, are identified very generally, placing no restriction on the serial dependence in the dependent variable. Furthermore these AMEs can be directly compared to each other, even though the response variable for CRE is binary and at the individual level whereas for the others it is fractional and at the cohort level, as well as the coefficients estimated using EWALD. Below we show how AMEs can be identified (based on Papke and Wooldridge, 2008) when using the Mundlak device with pseudo-panels. Although the derivation is based specifically on the fractional response models it also applies to CRE. We assume:

$$\mathrm{E}[\bar{y}_{ct} \mid \bar{\boldsymbol{x}}_{ct}, f_c] = \Phi(\bar{\boldsymbol{x}}_{ct}\boldsymbol{\beta} + f_c), \quad c=1,...,C \quad t=1,...,T \qquad (11)$$

Where $0 \le \bar{y}_{ct} \le 1$ and outcomes at the endpoints are allowed, in fact $\bar{y}_{ct}$ can even be binary. The independent variables, $\bar{\boldsymbol{x}}_{ct}$, are strictly exogenous and the unobserved heterogeneity, f_c, is additive inside the standard normal CDF by assumption. Inserting the conditional normality assumption following Mundlak (1978):

$$f_c \mid (\bar{\boldsymbol{x}}_{c1}, \bar{\boldsymbol{x}}_{c2}, ..., \bar{\boldsymbol{x}}_{cT}) \sim \text{Normal}(\psi + \bar{\bar{\boldsymbol{x}}}_c \xi, \sigma_a^2) \qquad (12)$$

Which can be rewritten as:

$$f_c = \psi + \bar{\bar{\boldsymbol{x}}}_c \xi + \mu_c \qquad (13)$$

Where $f_c \mid \bar{\boldsymbol{x}}_c \sim N(0, \sigma_a^2)$ and is independent of the explanatory variables, with $\bar{\bar{\boldsymbol{x}}}_c = \frac{1}{T}\sum_{t=1}^{T} \bar{\boldsymbol{x}}_{ct}$. If the regression includes any time dummies these are not included in $\bar{\bar{\boldsymbol{x}}}_c$.Substituting equation (13) into (11), we can see that *β*'s and the AMEs can be identified up to a positive scale factor:

$$\mathrm{E}[\bar{y}_{ct} \mid \bar{\boldsymbol{x}}_c, \mu_c] = \Phi(\psi + \bar{\boldsymbol{x}}_{ct}\boldsymbol{\beta} + \bar{\bar{\boldsymbol{x}}}_c \xi + \mu_c) \qquad (14)$$

Which can be rewritten as:

$$\begin{aligned} \mathrm{E}[\bar{y}_{ct} \mid \bar{\boldsymbol{x}}_c] &= \mathrm{E}[\Phi(\psi + \bar{\boldsymbol{x}}_{ct}\boldsymbol{\beta} + \bar{\bar{\boldsymbol{x}}}_c \xi + \mu_c) \mid \bar{\boldsymbol{x}}_c] \qquad (15) \\ &= \Phi[(\psi + \bar{\boldsymbol{x}}_{ct}\boldsymbol{\beta} + \bar{\bar{\boldsymbol{x}}}_c \xi)/(1+\sigma_a^2)^{1/2}] \end{aligned}$$

Or :

$$\mathrm{E}[\bar{y}_{ct} \mid \bar{\boldsymbol{x}}_c] = \Phi(\psi_a + \bar{\boldsymbol{x}}_{ct}\boldsymbol{\beta}_a + \bar{\bar{\boldsymbol{x}}}_c \xi_a) \qquad (16)$$

Where the *a* subscript denotes division by the factor $(1+\sigma_a^2)^{1/2}$. The scaled coefficients ψ_a, $\boldsymbol{\beta}_a$ and ξ_a are identified as long as there is no perfect linear relationship between the elements of $\bar{\boldsymbol{x}}_{ct}$ and there is some time variation in all the elements of $\bar{\boldsymbol{x}}_{ct}$, hence $\bar{\boldsymbol{x}}_{ct}$ and $\bar{\bar{\boldsymbol{x}}}_c$ are not perfectly collinear for all *t*. Estimating the above scaled model using PFP or GEE and using the law of large numbers we can consistently estimate the AMEs by either taking the derivatives or changes with respect to the explanatory variables. Thus if the regressor $\bar{x}_{ct1}$ is continuous its AME is:

$$\frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}\hat{\beta}_{a1}\phi(\hat{\psi}_a + \bar{\boldsymbol{x}}_{ct}\hat{\boldsymbol{\beta}}_a + \tilde{\bar{\boldsymbol{x}}}_c\hat{\xi}_a) \qquad (17)$$

To make the AMEs comparable to EWALD we need to use a single scale factor. This scale factor averages out $\bar{\boldsymbol{x}}_{ct}$ across all time periods and is given by:

$$\frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}\phi(\hat{\psi}_a + \bar{\boldsymbol{x}}_{ct}\hat{\boldsymbol{\beta}}_a + \tilde{\bar{\boldsymbol{x}}}_c\hat{\xi}_a) \qquad (18)$$

Dividing the fractional probit coefficient estimates ($\hat{\boldsymbol{\beta}}_a$) by this scale factor gives the AMEs which are directly comparable to the EWALD estimates. Note it is also possible to let the AMEs vary across time by allowing both the scale factor and AME calculations to change over time. If $\bar{x}_{ct1}$ is binary its AME can be estimated as:

$$\frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}[\Phi(\hat{\psi}_a + \hat{\beta}_{a1} + \bar{\boldsymbol{x}}_{ct(1)}\hat{\boldsymbol{\beta}}_{a(1)} + \tilde{\bar{\boldsymbol{x}}}_c\hat{\xi}_a) - \Phi(\hat{\psi}_a + \bar{\boldsymbol{x}}_{ct(1)}\hat{\boldsymbol{\beta}}_{a(1)} + \tilde{\bar{\boldsymbol{x}}}_c\hat{\xi}_a)] \qquad (19)$$

Where $\hat{\beta}_{a1}$ is the coefficient on $\bar{\boldsymbol{x}}_{ct1}$, $\bar{\boldsymbol{x}}_{ct(1)}$ are all explanatory variables except $\bar{\boldsymbol{x}}_{ct1}$ and similarly for its corresponding coefficient vector $\hat{\boldsymbol{\beta}}_{a(1)}$.

4 Monte Carlo Simulations

In order to test the small sample properties of the four estimates derived above we use Monte Carlo simulations based on the setup used by Collado (1998) to test the properties of their minimum distance based estimator. The main difference from their approach is that we base our results on the estimates of AMEs rather than coefficients. This approach allows the estimators to be tested under three different scenarios important for pseudo-panel applications: increasing autocorrelation in the explanatory variable, decreasing variation amongst the cohort population means, and increasing measurement error in the cohort sample means. Note all three factors are captured in

some way in the AWAR measure derived in the previous chapter for assessing the performance of pseudo-panel estimates.

The Monte Carlo setup considers a binary response model at the individual level with just one explanatory variable. The individual level data is generated by first assuming a cohort level relationship. Thus the data generating process involves first generating the underlying cohort population means for the explanatory variable (x^*_{ct}) using an AR[1] process where:

$$x^*_{ct} = \rho x^*_{c(t-1)} + w_{ct} \qquad (20)$$

The initial values of this series (x^*_{c0}) are generated as i.i.d. $N(0,\sigma^2_x)$ with $w_{ct} \sim N(0,\sigma^2_w)$ and the first ten cross-sections are discarded. This way the variance of the cohort population means can be calculated as:

$$\sigma^2_x = \frac{\sigma^2_w}{1 - \rho^2} \qquad (21)$$

This allows for the autocorrelation coefficient ρ to be varied while adjusting σ^2_w to keep σ^2_x constant. The unobserved heterogeneity for the cohort population is generated as:

$$\eta_c = x^*_{c1} + \ldots + x^*_{cT} + u_c \qquad (22)$$

Where u_c is i.i.d. $N(0,\sigma^2_u)$. Then the individual level regressors can be generated as:

$$x_{it} = x^*_{ct} + \varsigma_{it} \qquad \varsigma_{it} \sim \text{i.i.d.}\, N(0,\sigma^2_\varsigma) \qquad (23)$$

As σ^2_ς increases this causes the individual level values (x_{it}) to diverge further from the cohort population means (x^*_{ct}), resulting in the cohort sample means ($\bar{x}_{ct}$) containing greater measurement error when they are used as estimates of the population means. Thus σ^2_ς is used to represent the degree of measurement error in the cohort level data. The individual level latent dependent variables (y'_{it}) are generated as:

$$y'_{it} = x^{*}_{ct}\beta + \eta_c + f_t + v_{it} \qquad f_t \sim \text{i.i.d. } N(0,1) \text{ and } v_{it} \sim \text{i.i.d. } N(0,1) \qquad (24)$$

Where f_t is a time effect common to all individuals in the same time period. From this the observed individual level binary dependent variables can be obtained as:

$$y_{it} = \begin{cases} 1 \text{ if } y'_{it} > 0 \\ 0 \text{ otherwise} \end{cases} \qquad (25)$$

Following Collado we set the number of cohorts to 50, use five time periods and set the number of individuals per cohort (cell sizes) also at 50, giving us 2500 individuals in each time period. We first generate the x^{*}_{ct}, η_c and f_t values for the 50 cohorts in each time period, setting β =1, and then generate the individual level variables (x_{it} and y_{it}) 1000 times for each set of replications. Using the individual level data we apply the four estimators discussed in the previous section. Note the two fractional probit estimators (PFP and GEE) require standard errors to be estimated by bootstrapping for which we use 500 replications, hence these estimators in effect require 500,000 replications for each set. Each set of replications alter either the autocorrelation in the cohort population means (ρ), the level of variance in the cohort population means (σ_x^2), or the measurement error in the cohort sample means (σ_ς^2). As discussed in the previous chapter, based on Verbeek and Nijman (1992), the bias of the pseudo-panel estimates are expected to increase as the variance across the cohort population means decreases relative to the measurement error. The bias is also expected to increase with ρ as increasing the autocorrelation while keeping σ_x^2 constant reduces the degree of variation over time in the cohort means relative to the variation across cohorts, which Devereux (2007a) shows exacerbates the bias from measurement error. We also assess the estimators over different values of the AWAR measure developed in the previous chapter and check whether a similar threshold exists to that of the linear continuous response model.

Before proceeding with the results from the simulations it is important to discuss how they are to be presented. As we are interested in the accuracy of the AME estimates we need to compute the "true" AME against which they can be assessed. The "true"

AMEs can be derived from the underlying cohort population relationship as shown in equation (26), where the standard normal density function is represented by ϕ and we set $\sigma_u^2 = 1$. From the equation one can see that the AMEs depend on x_{ct}^*, thus each Monte Carlo replication which produces different x_{ct}^* will produce different AMEs. Therefore even within the set of a thousand replications the AMEs will differ around a certain mean. This mean will furthermore change as one alters ρ and σ_x^2. Hence, in order to compare the results across different sets of simulations, the AME estimates need to be scaled against the true AME values. We report the mean across the 1000 replications of the estimated AMEs relative to the "true" AME for each estimator. Thus a value of one implies the estimated AMEs are unbiased, below one implies they are downward biased and the reverse for a value of more than one. We also report the root-mean-squared-error (RMSE) for the AME estimates which are again scaled in a similar fashion by the "true" AMEs, hence it can be interpreted as an absolute percentage error measure. Finally, we include the 95% confidence interval coverage rate which reports the proportion of estimates whose 95% confidence interval includes the "true" AME value.

$$AME = \frac{\beta}{\left(1+\sigma_u^2\right)^{1/2}} \phi\left(\frac{x_{ct}^* \beta + \eta_c + f_t}{\left(1+\sigma_u^2\right)^{1/2}} \right) \qquad (26)$$

In Tables 4.1, 4.2 and 4.3 we show the results for varying the three parameters of interest; σ_x^2, ρ, and σ_ς^2 respectively. All four estimators have almost identical mean values irrespective of the degree of variance at the cohort level, autocorrelation or measurement error. Thus there is little to separate any of the four estimators in terms of their average level of biasedness. All estimates are downward biased, reflecting the attenuation caused by measurement error in the cohort sample means, with the bias decreasing with σ_x^2 and increasing with ρ and σ_ς^2. The RMSE shows that although all estimators have bias that is almost identical on average, the individual linear estimates tend to have a larger absolute degree of bias than the nonlinear estimates. This is true for all three nonlinear estimators for almost all values of the three parameters being varied. There is evidence that the linear estimator improves relative to the nonlinear ones as σ_ς^2 increases and σ_x^2 decreases, with the RMSEs converging albeit with the

linear RMSE remaining slightly larger. This indicates EWALD is relatively less affected by low cohort level variation or high measurement error. However this does not hold true for autocorrelation in the explanatory variables to which the nonlinear estimators are more robust. They have both a lower initial RMSE when there is no autocorrelation and the RMSE increases at a slower rate with ρ compared to EWALD.

Amongst the nonlinear estimators, CRE and PFP deliver AME estimates which are identical, although their standard errors do differ resulting in different coverage rates. However even this is due to PFP using bootstrap standard errors, which if applied to CRE would result in identical standard errors as well. Thus we can be confident that transitioning from a binary dependent variable at the individual level to a fractional one at the cohort level has little effect on the estimation process in addition to allowing for model extensions that require the dependent variable to be traceable over time, such as dynamic or differenced models. CRE/PFP estimates have the lowest RMSE in all scenarios and PFP produces the best coverage rates of all estimators. The CRE coverage rates are generally lower because they have smaller standard errors as a result of the estimation process artificially extending the sample size by keeping the dependent variable at the individual level. To avoid this problem it may be better for such estimation processes to use bootstrapping. EWALD usually has the lowest coverage rates although, similar to the RMSE, as σ_ς^2 increases and σ_x^2 decreases coverages rates for all estimators deteriorate and converge to similar levels indicating EWALD is relatively less effected by them. This is also partially true when ρ increases, with the exception of PFP whose coverage rates remain quite robust to increasing autocorrelation.

In Table 4.4 we assess the performance of the four estimators under different values of the AWAR measure constructed in the previous chapter. For a linear model it was shown in the previous chapter that a threshold of 0.5 was sufficient to produce reasonable pseudo-panel estimates (where the mean bias and RMSE were both less than 10 percent) when the data had the same number of time periods, cohorts, cell sizes and ρ=0.5. For the binary response model all four estimators do also have a mean bias that is less than 10 percent when AWAR is 0.5, however the RMSE is higher at around 12 per cent when it was 8.5 for the linear model. A higher threshold of at least

0.6 is more suitable, with RMSE being about 10 per cent and coverage rates reaching closer to 90 per cent. Amongst the four estimators CRE and PFP are the best across all values for AWAR as they have the lowest RMSE, although the mean values are again almost always identical for all estimators. Furthermore, PFP is again found to produce the best coverage rates. When AWAR is low, EWALD performs similarly to CRE/PFP however as AWAR rises the two nonlinear estimators improve at a greater rate, suggesting the improvements the nonlinear estimators provide are conditional on having high AWAR values.

Table 4.1: Monte Carlo Simulation results for varying σ_x^2

σ_x^2		EWALD	CRE	PFP	GEE
0.2	Mean	0.910	0.909	0.909	0.896
	RMSE	0.126	0.123	0.123	0.131
	95% Coverage	0.812	0.814	0.817	0.742
0.4	Mean	0.953	0.952	0.952	0.938
	RMSE	0.088	0.084	0.084	0.093
	95% Coverage	0.899	0.898	0.908	0.872
0.6	Mean	0.968	0.967	0.967	0.961
	RMSE	0.077	0.071	0.071	0.073
	95% Coverage	0.916	0.912	0.936	0.918
0.8	Mean	0.976	0.974	0.974	0.967
	RMSE	0.072	0.065	0.065	0.066
	95% Coverage	0.921	0.925	0.946	0.950
1	Mean	0.981	0.979	0.979	0.973
	RMSE	0.070	0.061	0.061	0.064
	95% Coverage	0.918	0.932	0.963	0.947

Note: We set ρ=0, σ_ς^2=1, C =50, cell size = 50, T=5 and β=1 for the above simulations with 1000 repetitions.

Table 4.2: Monte Carlo Simulation results for varying ρ

ρ		EWALD	CRE	PFP	GEE
0	Mean	0.981	0.979	0.979	0.973
	RMSE	0.070	0.061	0.061	0.064
	95% Coverage	0.918	0.932	0.963	0.947
0.2	Mean	0.978	0.976	0.976	0.970
	RMSE	0.076	0.066	0.066	0.068
	95% Coverage	0.919	0.928	0.966	0.957
0.4	Mean	0.975	0.973	0.973	0.966
	RMSE	0.087	0.074	0.074	0.073
	95% Coverage	0.907	0.916	0.966	0.957
0.6	Mean	0.968	0.966	0.966	0.954
	RMSE	0.104	0.085	0.085	0.087
	95% Coverage	0.914	0.907	0.953	0.938
0.8	Mean	0.945	0.942	0.942	0.937
	RMSE	0.140	0.113	0.113	0.110
	95% Coverage	0.883	0.884	0.940	0.924

Note: We set σ_x^2=1, σ_ς^2=1, C =50, cell size = 50, T=5 and β=1 for the above simulations with 1000 repetitions.

Table 4.3: Monte Carlo Simulation results for varying σ_ς^2

σ_ς^2		EWALD	CRE	PFP	GEE
0.5	Mean	0.991	0.989	0.989	0.982
	RMSE	0.068	0.058	0.058	0.061
	95% Coverage	0.927	0.942	0.968	0.963
1	Mean	0.981	0.979	0.979	0.973
	RMSE	0.070	0.061	0.061	0.064
	95% Coverage	0.918	0.932	0.963	0.947
2	Mean	0.962	0.960	0.960	0.956
	RMSE	0.077	0.070	0.070	0.073
	95% Coverage	0.889	0.887	0.928	0.920
3	Mean	0.944	0.942	0.942	0.936
	RMSE	0.088	0.082	0.082	0.085
	95% Coverage	0.835	0.823	0.880	0.840
4	Mean	0.927	0.925	0.925	0.923
	RMSE	0.100	0.095	0.095	0.097
	95% Coverage	0.772	0.726	0.795	0.767

Note: We set σ_x^2=1, ρ=0, C =50, cell size = 50, T=5 and β=1 for the above simulations with 1000 repetitions.

Table 4.4: Monte Carlo Simulation results for varying AWAR

AWAR		EWALD	CRE	PFP	GEE
0.2	Mean	0.652	0.651	0.651	0.641
	RMSE	0.380	0.379	0.379	0.382
	95% Coverage	0.338	0.291	0.299	0.152
0.4	Mean	0.882	0.881	0.881	0.871
	RMSE	0.159	0.154	0.154	0.156
	95% Coverage	0.766	0.754	0.764	0.677
0.5	Mean	0.921	0.920	0.920	0.905
	RMSE	0.126	0.119	0.119	0.125
	95% Coverage	0.842	0.830	0.845	0.785
0.6	Mean	0.943	0.942	0.942	0.936
	RMSE	0.109	0.100	0.100	0.099
	95% Coverage	0.883	0.879	0.903	0.884
0.8	Mean	0.967	0.966	0.966	0.960
	RMSE	0.096	0.082	0.082	0.082
	95% Coverage	0.915	0.907	0.953	0.941
1	Mean	0.979	0.977	0.977	0.971
	RMSE	0.094	0.074	0.074	0.072
	95% Coverage	0.913	0.925	0.975	0.971

Note: We vary σ_x^2 to set the value of AWAR, keeping ρ=0.5, σ_ς^2=1, C =50, cell size = 50, T=5 and β=1 for the above simulations with 1000 repetitions.

Finally, in Table 4.5 we compare using EWALD in a binary response pseudo-panel setting to using LPM in a panel setting. The panel simulations are setup in an identical fashion where we essentially estimate using the cohort level data; x_{ct}^*, η_c and f_t are generated exactly as before, however the binary dependent variable is now created at the cohort level as well. One can see this from equations (27) and (28) given below which are the panel corollaries of equations (24) and (25). We compare the pseudo-panel estimates to two different settings, one where each cross-section has 50 observations, which is equivalent to the number of cohorts in the pseudo-panel. The other uses 2500 observations in each time period as this compares directly to the number of individual level observations in the pseudo-panel, in effect allowing us to compare panel data to repeated cross-sections of the same dimensions.

$$y'_{ct} = x_{ct}^*\beta + \eta_c + f_t + v_{ct} \quad \text{where } v_{ct} \sim \text{i.i.d. } N(0,1) \qquad (26)$$

$$y_{ct} = \begin{cases} 1 \text{ if } y'_{ct} > 0 \\ 0 \text{ otherwise} \end{cases} \qquad (27)$$

Table 4.5: Monte Carlo simulations comparing Panel and Pseudo-panel estimates

ρ	σ_x^2		EWALD	Panel models using LPM	
				N=50	N=2500
0	0.2	Mean	0.910	1.001	0.999
		RMSE	0.126	0.302	0.042
		95% Coverage	0.812	0.944	0.949
0	0.4	Mean	0.953	0.994	1.000
		RMSE	0.088	0.245	0.033
		95% Coverage	0.899	0.941	0.941
0	0.6	Mean	0.968	0.990	1.000
		RMSE	0.077	0.221	0.030
		95% Coverage	0.916	0.933	0.936
0	0.8	Mean	0.976	0.995	1.000
		RMSE	0.072	0.209	0.028
		95% Coverage	0.921	0.928	0.936
0	1	Mean	0.981	0.997	1.000
		RMSE	0.070	0.201	0.027
		95% Coverage	0.918	0.924	0.924
0.2	1	Mean	0.978	0.998	1.000
		RMSE	0.076	0.225	0.031
		95% Coverage	0.919	0.918	0.921
0.4	1	Mean	0.975	0.996	0.997
		RMSE	0.087	0.252	0.036
		95% Coverage	0.907	0.926	0.928
0.6	1	Mean	0.968	1.001	0.995
		RMSE	0.104	0.301	0.044
		95% Coverage	0.914	0.937	0.919
0.8	1	Mean	0.945	1.001	0.993
		RMSE	0.140	0.415	0.059
		95% Coverage	0.883	0.933	0.934

Note: We set σ_ς^2 =1, C =50, cell size = 50 for the pseudo-panel estimates, for all estimates we set T=5 and β=1 in the above simulations with 1000 repetitions

The LPM in both settings (with 50 or 2500 observations) produces estimates which are almost completely unbiased in all scenarios compared to EWALD where the estimates are always somewhat attenuated. This attenuation bias increases with higher autocorrelation and lower variance in the independent variable. However the bias is often not too severe being less than 10 percent even when ρ is large and σ_x^2 is small. Furthermore, EWALD has RMSE that are often two to three times smaller than those for the LPM model with 50 observations although they are also two to three times larger than those for the model with 2500 observations. Thus EWALD can be a better estimator than its panel counterpart when the panel contains the same number of

observations as at the cohort level in the pseudo-panel, particularly for low values of ρ and high values of σ_x^2. There is a trade-off between the unbiased LPM which produces estimates that are on average correct (and also have better coverage rates) and EWALD which has lower RMSE meaning the absolute size of the bias of each individual estimate is likely to be a lot smaller. In contrast, EWALD is unambiguously the poorer estimator compared to a panel model which has the same number of individual observations.

The results from the Monte Carlo simulations indicate that EWALD can be used as a suitable pseudo-panel extension of the LPM used with panel data. Naturally attenuation bias is present in EWALD as the cohort sample means are estimates of the underlying cohort population while no such bias arises in panel data models. However as long as autocorrelation or measurement error is not too high nor cohort level variance too low (this can be expressed as AWAR being at least 0.5 or 0.6) the bias is not very large and coverage rates are reasonable. The RMSE are also quite moderate as well as being much smaller than those of the panel counterpart which has the same dimensions as at the cohort level. The large difference in RMSE between panel estimates with 50 and 2500 cross-section observations show that when constructing cohorts for binary response pseudo-panel models one may be better served by expanding the number of cohort level observations even if it increases the attenuation bias by reducing the cell sizes.

EWALD can produce estimates that are just as accurate for a binary response model as those for a linear model but with stricter requirements on cohort construction indicated by the higher AWAR threshold. The nonlinear estimators have been shown to offer improvements to EWALD in terms of reduced RMSE and better coverage rates, although all four estimators equally suffer from attenuation bias. When cohorts are constructed well (i.e. AWAR is quite large) or there is a high degree of autocorrelation then CRE and PFP are preferable to use. EWALD does appear to be relatively less affected by poor cohort construction (factors that lead to a low AWAR), however even in these scenarios CRE and PFP are generally no worse than EWALD. Furthermore, even when using the nonlinear estimators a higher threshold for AWAR is still required than in the linear case.

5 Empirical Application

Results from the Monte Carlo simulation give cause for optimism that all four binary response pseudo-panel estimators can deliver accurate estimates of AMEs albeit with slightly stricter conditions than is required for a linear model. We now test if this holds empirically by using panel data to compare panel and pseudo-panel estimates in a similar fashion to the previous chapter, which showed that as long as an AWAR threshold was met panel and pseudo-panel estimates were very similar for the linear fixed effects model. Using the same dataset, four waves of the Ugandan National Panel Survey, we estimate a direct binary response model extension of the linear household consumption model used in the previous chapter.

We alter the model slightly from one assessing the determinants of household consumption to that of household poverty, where we define a household as being poor if its adult equivalent consumption per head is below the median level for all households in the same time period. The median threshold will overstate poverty compared to those based on a national poverty line or to the dollar-a-day ($1.90) measure. Nevertheless we use this threshold due to its simplicity and the convenience of splitting the sample evenly. Furthermore it is quite a reasonable measure when compared to the higher $3.10 a day poverty line. The alteration of the model in this way serves a twofold purpose; firstly allowing for the direct comparison of linear and nonlinear pseudo-panel estimation using essentially the same model and data. Secondly it offers a direct application of how pseudo-panel estimation can be used in the analysis of poverty at a more disaggregated level to complement the country-level analysis conducted in the first chapter. It should be noted that our binary dependent variable is zero if the household is defined as poor (below the median consumption level) and one if not, thus the model looks into factors determining households being out of poverty. We estimate in this way so the coefficients are directly comparable to the household consumption model, making it easier to compare the effects of transitioning from a linear to a binary response model. The model to be estimated in given as:

$$\begin{aligned} \text{bcons}_{it} &= \alpha + \beta_1 \text{HHsize} + \beta_2 \text{ltotasset} + \beta_3 \text{secondary} + \beta_4 \text{lnremit} \\ &\quad + \beta_5 \text{remittance} + \boldsymbol{\delta z}_{it} + f_i + f_t + e_{it} \end{aligned} \tag{28}$$

where *bcons* is binary dependent variable indicating if consumption if above the median level, *HHsize* is the size of the household, *ltotasset* is the log of total assets owned by the household not including agricultural and enterprise assets, *lnremit* is the log of remittances received by the household in the past year, *remittance* is a dummy for whether the household received any remittance in the past year, *secondary* is a dummy variable indicating whether the household head had a secondary source of income in the past year, $\boldsymbol{z}$ is a vector of other controls, f_i indicates the household specific effects potentially correlated with some of the regressors and f_t represents time effects.

In Table 4.6 we estimate this model using both linear and probit estimation making use of the panel nature of the data. Columns 1 and 3 estimate without accounting for the potential presence of fixed effects, although additional controls are included. These controls are the variables used to construct cohorts when using pseudo-panel estimation; namely the age of the household head (and its squared term), the region the household is from, whether it is located in an urban or rural area, the gender and education level of the household head. Columns 2 and 4 do allow for fixed effects by using a fixed effects specification for the linear estimator and CRE (the Mundlak version which includes the mean across time for each regressor) for probit, however the additional controls are omitted. The Breusch and Pagan Lagrangian multiplier test and the robust version of the Hausman test imply fixed effects are present and need to be controlled for. The pseudo-panel estimates will be assessed against the estimates reported in Table 4.6. We assume that the fixed effects and CRE estimates produce the most accurate AMEs, thus the closer pseudo-panel estimates are to these the better they are assume to be. The OLS and probit estimates serve to show if pseudo-panel estimation offers an improvement over using cross-sectional models which instead use the variables required for cohort construction as additional controls, where an improvement is classed as having AME estimates that are closer to FE or CRE. We also trim the data at the 1st and 99th percentiles to remove the effects of outliers as they would be expected to have less impact on pseudo-panel estimates.

Table 4.6: Panel estimates of Households having above Median Consumption

VARIABLES	OLS	FE	Probit	CRE
HHsize	-0.054***	-0.050***	-0.054***	-0.053***
	(0.002)	(0.003)	(0.002)	(0.003)
ltotasset	0.091***	0.056***	0.095***	0.072***
	(0.003)	(0.005)	(0.003)	(0.006)
secondary	0.006	0.032***	0.001	0.035***
	(0.009)	(0.011)	(0.009)	(0.010)
lnremit	0.035***	0.015**	0.038***	0.019***
	(0.005)	(0.006)	(0.005)	(0.007)
remittance	-0.393***	-0.173**	-0.426***	-0.211***
	(0.060)	(0.071)	(0.063)	(0.064)
Other controls	Yes		Yes	
Observations	9,510	9,510	9,510	9,510

Note: Robust standard errors in parentheses. The probit and CRE columns show the estimates of the average marginal effects. All regressions contain year dummies to allow for year fixed effects. Other controls refer to the variables used to construct cohorts; age, gender, and education of household head, the region the household is from and whether it is in an urban or rural location. All regressions trim at the top and bottom one percentiles to limit the impact of outliers.

The results in Table 4.6 are consistent with those from the linear model in the previous chapter, all having the same sign and significance levels although naturally the estimates are of different sizes given one set are AME whereas the others are coefficients. Allowing for the presence of fixed effects also has the same effect as before for all regressors: *HHsize* is largely unaffected; *secondary* becomes positive and significant whereas before it was insignificant; *ltotasset*, *lnremit* and *remittance* keep the same sign and significance but the coefficients fall in absolute size. The FE and CRE estimates are largely identical except for *ltotasset* where the CRE estimate is somewhat larger. We compare these to the pseudo-panel estimates shown in Tables 4.7, 4.8, and 4.9. Each table contains the four estimators discussed above (we rename CRE for the pseudo-panels as PP CRE to avoid confusion with the panel CRE results in Table 4.6) with an additional estimator in the first column named EWALD H. This estimator is the same as EWALD but like with PP CRE we keep the binary dependent variable at the household level. The three tables differ in how cohorts were constructed: Table 4.7 uses two-year age cohorts combined with four of the other construction variables - region of household, whether it is in an urban or rural location, gender and education level of the household head. Table 4.8 is similar but does not

include gender in the cohort specification while Table 4.9 excludes the urban/rural location instead. In all three tables we include the AWAR value for each regressor. We pick these three cohort construction specifications as they have been shown to offer the best estimates for the linear model in the previous chapter and produce the highest AWAR values. Thus they represent the best estimates that can be produced from the dataset available.

Across all three pseudo-panel results the estimates for *HHsize* are largely very similar in terms of sign, significance and size. No one estimator consistently outperforms the others and all are very similar to the panel estimates, if slightly smaller in terms of size. There is some evidence that pseudo-panel estimates improve as AWAR increases, although across the three tables there is little variation in the values for *HHsize*. Similar results hold for *ltotassets*, with all pseudo-panel estimates being very close to the panel ones. In fact the pseudo-panel estimates also somewhat mimic the nonlinear CRE estimate being of a larger size than the linear FE estimate. The estimates for *secondary* are generally poorer, with only EWALD H being able to produce positive and significant estimates similar to FE or CRE. EWALD produces estimates almost identical in size to EWALD H, however their higher standard errors lead to the AME being insignificant. For the other three nonlinear estimators it is a combination of larger standard errors and smaller AME which leads to their lack of significance. Amongst the three nonlinear estimators PP CRE is the better estimator, although all three are quite poor but they do offer some improvement to the cross-sectional estimates.

The coefficients on *lnremit* and *remittance* need to be analysed together as their effects are inter-related, evidence of this can be seen in Table 4.6 where although the absolute size of the estimates change their size relative to each other remains roughly constant. Assessing which estimator is better than the others is slightly harder for these two regressors. EWALD and EWALD H produce AMEs which are the closest to the panel estimates in terms of size, although EWALD is by far the worst of the five estimators in terms of significance. The nonlinear estimators are generally better than even EWALD H in terms of significance levels for *lnremit* and *remittance*, however the AME estimates are too large and in some cases worse than the cross-sectional

estimates although not to a large extent. Amongst the three nonlinear estimators GEE is the best in terms of estimating AME size but PFP is best for significance levels.

The Monte Carlo simulations displayed results suggesting that pseudo-panel estimators for a binary response model would require higher AWAR values than for a linear model. We can see evidence of this in our empirical application by comparing the results from the binary response model estimated above to the linear model in the previous chapter. On the whole the pseudo-panel estimators in the binary case are less proficient at producing estimates similar to their panel counterparts than in the linear case. Even for *HHsize* and *ltotassets*, which are generally done quite well in the binary case, the linear model has pseudo-panel estimates which are closer in size to their panel counterparts (although performance is equal in terms of significance levels). This is true more so for *HHsize* than *ltotassets* and this could be because while an AWAR value of 0.5 is fine in the linear case, the simulations in the previous section imply a value of at least 0.6 is required for binary response models. *HHsize* is not able to meet this higher threshold but *ltotassets* is. Failure to meet the higher threshold also occurs for *secondary* and in combination with the fact that it is a binary variable, which even in the linear model makes it relatively more difficult to estimate using pseudo-panels, may mean an even higher threshold is required thus explaining why it is so poorly estimated. As the simulations implied that EWALD was relatively more robust to low AWAR values than the nonlinear estimates, this may be why the two linear estimators are better at estimating the AME for *secondary*. Even then EWALD H needs to be used to ensure the estimates are significant. For *lnremit* and *remittance* AWAR values are higher but EWALD H is still on balance the best estimator, suggesting that for regressors with some element of discreteness linear estimators may be more robust. Even for a linear model, pseudo-panel estimates are less proficient at estimating such regressors; the problem is exacerbated for binary response models especially when using nonlinear estimators. The nonlinear estimates can deliver correct significance levels for these variables but again the results suggest that a higher AWAR threshold of at least 0.6 is required for this. This higher threshold may be too difficult to obtain when using some datasets. For example, in our data *HHsize* and *secondary* struggle to meet this threshold even when we use some of the best possible specifications.

Overall EWALD H produces AME estimates that are the most reflective of the panel results especially when compared to its linear counterpart, the FE estimates. This seems to be due to it being more robust to problems arising from either low AWAR values or some element of discreteness in the variable or a combination of the two. In comparison, the nonlinear pseudo-panel estimates are not as good at replicating the CRE estimates. For the two remittance variables there is some evidence in favour of using nonlinear estimates as they produce estimates with better significance levels but at the expense of overestimating the AME sizes. Also there may be some benefits for estimating *ltotassets* with a nonlinear estimator as they produce estimates that are a little closer to the CRE estimate. EWALD H does require using the dependent variable at the household level and if this is not a suitable option then there is no clear estimator which would be preferable: PFP and GEE are better for achieving the correct significance of estimates but EWALD is better for estimating their size. One potential way to address the lack of significance for EWALD is to estimate using the within transformation rather than cohort dummies. This does improve the efficiency of estimates but for our model this improvement generally is not sufficient to improve significance levels. We also find that PP CRE and PFP no longer produce identical AME estimates, although they are generally quite similar and results are largely unchanged whether one uses the one or the other. Therefore changing the dependent variable from being at the household level to the cohort level has little impact, at least for nonlinear estimators. Using GEE instead of PFP also has little impact on the results.

The three nonlinear estimators offer improvements to using a cross-sectional model even if they fall short of exactly replicating the panel results. The estimates do retain the correct significance levels, especially when AWAR values are larger, and coefficient sizes are closer to the panel values than the cross-sectional estimates, particularly for *ltotassets* and *secondary* even though the latter fails to become significant. The linear estimators are very adept at producing AME size estimates that are similar to the panel ones, offering large improvements to the cross-sectional estimates. However they, EWALD in particular, struggle to maintain significance levels for *lnremit* and *remittance*, which makes them less reliable than the corresponding cross-sectional estimates. Whereas the problems with *secondary* do

appear to originate from low AWAR values, this is not the case with the two remittances variables. The linear estimators if anything have worse significance levels as AWAR increases, suggesting there may be other problems when estimating such variables. One concern may be the fact the two variables interact with one another and have effects going in opposite directions.

Table 4.7: Pseudo-panels using 2-year age cohorts, region, location, gender and education

Variables (AWAR)	EWALD H	EWALD	PP CRE	PFP	GEE
HHsize (0.55)	-0.048***	-0.048***	-0.047***	-0.044***	-0.044***
	(0.007)	(0.008)	(0.008)	(0.006)	(0.006)
ltotasset (0.63)	0.051***	0.051***	0.063***	0.054***	0.050***
	(0.008)	(0.010)	(0.009)	(0.010)	(0.009)
secondary (0.60)	0.034**	0.034	0.027	0.013	0.019
	(0.013)	(0.032)	(0.026)	(0.025)	(0.025)
lnremit (0.75)	0.020*	0.020	0.041***	0.039***	0.031**
	(0.011)	(0.013)	(0.012)	(0.014)	(0.013)
Remittance (0.78)	-0.238	-0.237	-0.498***	-0.452***	-0.359**
	(0.148)	(0.158)	(0.147)	(0.168)	(0.150)
Observations	8,798	1,741	8,798	1,741	1,741

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4.8: Pseudo-panels using 2-year age cohorts, region, location and education

Variables (AWAR)	EWALD H	EWALD	PP CRE	PFP	GEE
HHsize (0.50)	-0.040***	-0.038***	-0.043***	-0.045***	-0.041***
	(0.003)	(0.007)	(0.005)	(0.007)	(0.007)
ltotasset (0.60)	0.057***	0.057***	0.081***	0.079***	0.067***
	(0.005)	(0.012)	(0.009)	(0.014)	(0.012)
secondary (0.51)	0.021	0.021	0.017	-0.008	0.008
	(0.030)	(0.032)	(0.022)	(0.029)	(0.028)
lnremit (0.57)	0.023**	0.022	0.033*	0.034**	0.027*
	(0.011)	(0.015)	(0.019)	(0.017)	(0.015)
Remittance (0.58)	-0.262*	-0.247	-0.379	-0.380*	-0.301*
	(0.140)	(0.181)	(0.239)	(0.206)	(0.172)
Observations	9,489	1,450	9,489	1,450	1,450

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4.9: Pseudo-panels using 2-year age cohorts, region, gender and education

Variables (AWAR)	EWALD H	EWALD	PP CRE	PFP	GEE
HHsize (0.52)	-0.042***	-0.042***	-0.044***	-0.042***	-0.042***
	(0.005)	(0.007)	(0.005)	(0.006)	(0.006)
ltotasset (0.60)	0.048***	0.048***	0.059***	0.052***	0.049***
	(0.012)	(0.011)	(0.010)	(0.011)	(0.010)
secondary (0.53)	0.056***	0.053	0.045	0.034	0.037
	(0.014)	(0.035)	(0.032)	(0.029)	(0.028)
lnremit (0.66)	0.024**	0.024*	0.042***	0.039**	0.033**
	(0.009)	(0.014)	(0.013)	(0.015)	(0.014)
Remittance (0.68)	-0.267**	-0.267	-0.475***	-0.422**	-0.373**
	(0.124)	(0.171)	(0.164)	(0.180)	(0.165)
Observations	9,431	1,484	9,431	1,484	1,484

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

One concern with the empirical application is that the results are being driven by potentially endogenous variables (particularly those linked to income i.e. secondary occupation and the remittances variables) and as a consequence the results above are being driven by how the estimators deal with endogeneity rather than the pseudo-panel nature of the data. To address this concern we re-estimate model using a simplified version that excludes secondary occupation and the remittance variables. The results are presented in the Appendix in Tables 4.A1-4.A4, which replicates the empirical application presented in Tables 4.6- 4.9 as a robustness check. The results indicate that dropping potentially endogenous variables has little impact on the main findings, with the two EWALD estimates being marginally better than the other estimators at reproducing the genuine panel estimates across the three pseudo-panel specifications.

6 Allowing for Endogeneity

The fixed effects model we have considered thus far, both in this chapter and the previous, allows for the error term to be correlated with explanatory variables assuming the correlation is due to time invariant unobservables. If some of the unobservables are also time varying than fixed effects alone cannot produce consistent

estimates as the explanatory variables suffer from endogeneity. Without addressing concerns regarding endogeneity it is difficult to identify causality from the estimated model. However finding valid instruments to address endogeneity concerns is often difficult to do in practice. This section discusses how the pseudo-panel estimators can be extended to allow for both fixed effects and endogeneity. Then, making use of the ability to track variables over time in a pseudo-panel, we consider the potential benefits of a simple solution commonly used for tackling endogeneity where lagged variables are used as instruments. Empirical application and testing of these methods for incorporating endogeneity into pseudo-panel models are left for future work.

The EWALD estimator is the simplest with which one can incorporate instrumental variables into the analysis. Two-stage least squares estimation can be conducted either using cohort dummies or the within transformation using the cohort level data. Deaton (1985) derives an IV estimator which also includes a correction for measurement error caused by small cell size which leads to the cohort sample means being poor estimates of the cohort population means. Therefore if one is struggling to create enough cohort observations so that there is sufficient variation at the cohort level while maintaining reasonable cell sizes, the measurement error corrected estimator may be a useful alternative as it will reduce the cell sizes required. For binary response models, estimating a LPM using 2SLS is likely to give good estimates of average marginal effects (Wooldridge, 2010a) and one would expect this to also hold for estimating such models with EWALD.

It is generally more difficult to incorporate instrumental variables into a nonlinear model, such as the CRE and fraction response models. 2SLS is generally inconsistent for nonlinear models both for estimating parameters and average marginal effects (Wooldridge, 2014). Instead a two-step control function approach, where residuals from a first stage regression for the endogenous variables are inserted into the second stage estimation, can be used. Rivers and Vuong (1988) and Wooldridge (2010a) apply this approach to probit models. Blundell and Powell (2003, 2004) show that this approach can be used to identify the marginal effects very generally. However, the main drawback of the control function approach is that it can only be used for continuous endogenous explanatory variables and not for discrete ones. Terza et. al.

(2008) and Wooldridge (2014) modify the control function approach in different ways so the application can be extended to discrete variables as well. For our purposes Wooldridge's extension is more useful as it focuses on estimating AMEs rather than parameter coefficients. An application of Wooldridge's (2014) approach can be found in Papke and Wooldridge (2008) for their fractional response model with a continuous endogenous regressor and correlated random effects. Estimation is fairly simple to do using quasi-MLE however it is inconsistent using a GEE approach. Thus the PFP and CRE estimators can incorporate instrumental variables using this approach but not GEE. Furthermore, results from Papke and Wooldridge show that the control function approach produces estimates which are very similar to a linear IV estimator, indicating the potential for the linear EWALD estimator to be used for such models as well.

So far we have discussed how three of our four proposed estimators can be extended to incorporate endogeneity. If a valid instrument is available the advantage to these pseudo-panel estimators over simple OLS when one has RCS data is that the former can allow for fixed effects and endogeneity whereas the latter only allows for endogeneity. In practice finding a valid instrument which satisfies both the exogeneity and relevance criteria in order to identify casual inferences can be quite difficult. However, pseudo-panel estimation may offer a simple way to find such an instrument that cannot be done otherwise when using RCS data. For panel or time-series data the lags of a regressor can be used as instruments for itself as long there is sufficient correlation between the regressor and its lags and the lags do not have a direct effect on the independent variable. This approach can be extended to pseudo-panel estimation as the cohort level observations can be traced through time, allowing their lags to be used as instruments in such a way, even though the individual/household level observations are not traceable over time.

Reed (2015) looks into one specific cause of endogeneity, due to simultaneity, and shows simply replacing an endogenous regressor with its lag does not reduce the simultaneity bias. However, using lagged values as instruments (by using 2SLS, GMM or LIML) can be an effective estimation strategy as long as the lagged values are sufficiently correlated with contemporaneous values and the lags do not themselves belong in the estimating equation. Bellemare, Masaki and Pepinsky (2015) focus on

more general forms of endogeneity and assess whether lagged values can estimate causal effects. Their analysis indicate that even for the simultaneity problem an additional condition to those stated by Reed is required; one needs to also assume that there are no dynamics (no serial correlation) in the independent variable. They also find that for more general form of endogeneity the addition condition is that there are no dynamics in the error term. The intuition is that as endogeneity is due to correlation between a regressor and the error term, although a lagged regressor may not be correlated with a contemporaneous error term it will be correlated with a lagged error term. Thus for it to be a valid instrument there must be no serial correlation in the error term.

If one can assume that the unobserveables, which are correlated with the regressors and time varying, are not serially correlated then pseudo-panel estimation can be used to find simple instruments to address endogeneity concerns. This assumption may be difficult to defend in many applications particularly as it cannot be tested given it is an assumption regarding unobserveables. Bellemare, Masaki and Pepinsky (2015) highlight some cases where the assumption can be relaxed and the use of lagged variables can be justified. Hence pseudo-panel estimators like EWALD, CRE and PFP (which can be extended to incorporate instrumental variables) are not just useful for incorporating fixed effects but also for addressing endogeneity concerns especially when valid external instruments may be difficult to find. The testing of such extensions and whether it affects the required AWAR or CAWAR threshold is beyond the scope of this thesis and is left for future work.

7 Conclusion

This chapter has proposed and assessed the performance of several estimators which can be applied to a binary response model where unobserved heterogeneity may be correlated with regressors and the data is a series of independent repeated cross-sections. We consider the application of the linear pseudo-panel estimator, EWALD, to such a setting and propose three alternative nonlinear estimation techniques based

on correlated random effects models. We show how the AMEs of the nonlinear estimators can be derived and thus compared to EWALD coefficients. Then we assess the performance of these estimators using Monte Carlo simulations and an empirical example where we compare the pseudo-panel results to ones from a genuine panel.

Both the simulations and empirical application indicate that pseudo-panel estimates can be applied to binary response models just as well as for linear models, however they require stricter conditions to achieve estimates of similar reliability. Specifically, the results show that an AWAR threshold of at least 0.6 is required for binary response models whereas a threshold of 0.5 is sufficient for linear models. This means the requirements on cohort construction is even more severe for binary response models, to such an extent that some datasets may not be able to meet the higher threshold. For example, in our empirical application we were not able to construct cohorts in a way that this threshold could be met for some of our variables.

Our results show that EWALD, especially when estimated keeping the dependent variable at the individual level, is proficient at estimating binary response models in a not too dissimilar fashion to the LPM with panel data. There is evidence that EWALD is more robust to problems arising from cohort construction, where AWAR values are low because either measurement error is high or cohort-level variance is low, than the nonlinear estimators particularly when estimating the size of AMEs. Thus if one is struggling to construct cohorts with large AWAR values EWALD is likely to be the better option. Conversely, when AWAR values are large, especially if there is a high degree of autocorrelation in the regressor, then nonlinear estimators may be better. Of these the two fractional response estimators perform just as well as the simplified version of Collado's (1998) CRE estimator, indicating there is no loss in estimation performance caused by moving from using the dependent variable at the individual/household level to the cohort level. However, this does not hold for EWALD where estimating with the dependent variable at the individual level does improve results particularly for determining the significance level of a variable.

Regressors that are binary, censored or have some type of interaction with another variable are more difficult to estimate and replicate panel results using pseudo-panels.

The analysis does show that models requiring inclusion of such regressors also require higher AWAR thresholds. We have used the AWAR thresholds because they are easier to compare with results from the linear model in the previous chapter. However what is more relevant for applied work is to calculate the CAWAR (cell size adjusted AWAR) thresholds that can be used with repeated cross-sections for binary response models. These thresholds may also differ as the number of cohort level observations changes, as suggested by the large difference in RMSE between the panel simulations with 50 and 2500 cross-sectional observations. Furthermore, different thresholds may also be required when estimation is extended to allow for endogenous regressors.

Appendix to Chapter 4

Table 4.A1: Re-estimation of Table 4.6 with the Simplified Model

VARIABLES	OLS	FE	Probit	CRE
HHsize	-0.053***	-0.049***	-0.054***	-0.053***
	(0.002)	(0.003)	(0.002)	(0.003)
ltotasset	0.094***	0.057***	0.098***	0.074***
	(0.003)	(0.005)	(0.003)	(0.006)
Other controls	Yes		Yes	
Observations	9,533	9,533	9,533	9,533

Note: As for Table 4.6

Table 4.A2: Re-estimation of Table 4.7 with the Simplified Model

Variables (AWAR)	EWALD H	EWALD	PP CRE	PFP	GEE
HHsize (0.55)	-0.047***	-0.047***	-0.045***	-0.045***	-0.046***
	(0.007)	(0.008)	(0.010)	(0.007)	(0.006)
ltotasset (0.63)	0.052***	0.052***	0.062***	0.056***	0.051***
	(0.008)	(0.010)	(0.011)	(0.010)	(0.009)
Observations	8,798	1,838	8,798	1,838	1,838

Note: As for Table 4.7

Table 4.A3: Re-estimation of Table 4.8 with the Simplified Model

Variables (AWAR)	EWALD H	EWALD	PP CRE	PFP	GEE
HHsize (0.50)	-0.040***	-0.038***	-0.042***	-0.046***	-0.041***
	(0.003)	(0.007)	(0.006)	(0.007)	(0.007)
ltotasset (0.60)	0.059***	0.058***	0.086***	0.079***	0.065***
	(0.005)	(0.012)	(0.009)	(0.016)	(0.013)
Observations	9,489	1,450	9,489	1,450	1,450

Note: As for Table 4.8

Table 4.A4: Re-estimation of Table 4.9 with the Simplified Model

Variables (AWAR)	EWALD H	EWALD	PP CRE	PFP	GEE
HHsize (0.52)	-0.041***	-0.041***	-0.041***	-0.040***	-0.040***
	(0.005)	(0.007)	(0.006)	(0.006)	(0.006)
ltotasset (0.60)	0.049***	0.049***	0.058***	0.054***	0.051***
	(0.011)	(0.011)	(0.009)	(0.011)	(0.010)
Observations	9,431	1,484	9,431	1,484	1,484

Note: As for Table 4.9

Chapter 5:

Conclusions

This thesis covers a diverse set of approaches which are used to analyse poverty at different levels of aggregation. The first study, which uses country level data, highlights the large variation in poverty reduction experience across countries in Sub-Saharan Africa, linking it to a nation's colonial heritage and its political commitment to reducing poverty. The second and third studies investigate how the wealth of national household surveys that have been collected over recent decades across many SSA countries can be used to analyse poverty dynamics and related topics. Both approaches have the potential to offer important insights into different aspect of poverty but both also pose certain methodological and data difficulties in their applications.

In the first study (Chapter 2) we provide evidence suggesting that poverty dynamics across Africa since the 1990s is rooted in a country's colonial history, particularly the structure of its colonial economy. This economic structure is used to categories countries into two groups; crop-based and resourced-based. Crop-based countries were those that followed a colonial strategy of peasant export where infrastructure was developed for exporting smallholder cash crops. This fostered broad-based development with relatively strong linkages between agriculture and the rest of the economy. Resource-based countries were those that predominately had a history of settler capitalism where European settlers were allowed to appropriate land for the production of export crops (often plantations) or mining. These countries are usually associated with relatively high inequality, capital intensity and a narrow economic structure with weak linkages across sectors.

The empirical evidence strongly suggests that crop-based economies have been more successful at reducing poverty than resource-based ones. Growth is also significantly associated with poverty reduction only in crop-based countries whereas for resource-based growth and poverty are positively related. This hold true when different poverty

lines are used (both national and international), the effect of other covariates are controlled for, and different estimation strategies are utilised. The findings suggest that poverty has on average halved in the crop-based economies while there has been no significant reduction for the resource-based. Other results indicate that democracy and lower inequality are associated with higher poverty for resource-based economies, while greater pro-poor expenditure by governments (a proxy for redistributive policies) does permit lower poverty but again only in resource-based economies whereas they are not instrumental for poverty reduction in crop-based ones.

Due to the nature of the data available the study is limited to simply being able to test if there has been a difference between crop and resource-based economies in their poverty reduction performance rather than explain poverty reduction itself. The number of observations are limited due to the necessity of using survey data to produce reliable poverty estimates, which occur quite infrequently in many African countries. Even if more frequent data were available, when one is trying to explain poverty reduction performance over a period of time there are only as many genuine observations as there are countries. In our context that would mean there are around 30 genuine observations with which meaningful analysis can be conducted. This makes analysing poverty dynamics at the country level quite difficult when focusing just on Sub-Saharan Africa. Thus one is limited in terms of the estimation strategies that can be used and the inferences that can be drawn, with results being suggestive rather than allowing for causal inference. This is particularly true when analysing some of the main covariates of interest such as inequality, democracy, and pro-poor expenditure and how they are linked to a country's colonial legacy. Nevertheless, the analysis does show that incorporating the impact of smallholder interests, particularly in the political process, on poverty reduction remains an important area for future research.

The other two studies discuss a possible solution to one of the main drawbacks to analysing key issues in African economies, such as poverty, at the household or individual level. Over the last few decades a lot of data have been collected at this level in the form of consumption, agricultural, demographic, and health surveys. Due to the great cost and difficulty of re-interviewing the same set of individuals

particularly over the medium and long term (due to attrition, migration, family break-up), most of the data is in the form of repeated cross sections (RCS) where each wave contains a different set of individuals. When there is panel data available these tend to be over a relatively short period of time which often is insufficient for a meaningful analysis of changes and dynamics occurring in a country. RCS data also generally have a larger sample size hence are more representative of the overall population, whereas panel data may only be available for certain regions within a country. However, even with the longer coverage and more representative sampling, the available RCS data may still be inadequate for addressing many pertinent issues because individual are not traceable over time. For example, one key area of research regarding poverty of interest to policymakers is finding out the factors which help move households out of poverty and the main shocks or vulnerabilities that may lead to them slipping back into it. Such analysis can only be done with panel data where a household can be traced over the different surveys to track their movements in and out of poverty. This limitation is not just restricted to differenced models of this type but any model that requires a dynamic specification such as lagged dependant variables. It also precludes the possibility of allowing for fixed effects in the model which is necessary for consistent estimation if there are time invariant unobservables that are correlated with the regressors. If there is also possible endogeneity issues then RCS data excludes the possibility of a simple solution in the form of using lagged values as instruments.

The solution to this discussed in the thesis proposes using pseudo-panel estimation on RCS data. This involves grouping households or individuals into cohorts based on time invariant characteristics such as birth cohort or gender. Then a synthetic panel can be created using the means of the cohorts which are now traceable over time. Essentially the cohort means are used as if they are a representative household/individual with which one can carry out panel-type estimation. Chapter 3 shows the critical importance of cohort construction for pseudo-panel estimation. It highlights that the way cohorts are created affects how representative the cohort means are of the household level data. This in turn affects the consistency of the pseudo-panel estimates and whether they are similar to estimates derived if genuine panel data had been available. The chapter finds that many of the common methods of forming

cohorts discussed in the literature are inadequate as the pseudo-panel estimates are biased and often worse than if OLS was used. However other methods of constructing cohorts are found to produce estimates that are almost identical to panel estimates. We then propose a measure to test whether a cohort construction method is likely to produce reliable estimates. The efficacy of this measure is shown using an empirical example and Monte Carlo simulations.

The analysis is based on a linear fixed effects model and therefore the empirical application is conducted on a model of household consumption. Although these models can give important insights into the behaviour and changing circumstances of households they are insufficient for the analysis of poverty. Models looking into the determinants of poverty, or even movements in and out of poverty, have a binary dependent variable and as such the linear estimation used in Chapter 3 may not be suitable as it does not account for the boundedness of the response variable. This matter is explored in Chapter 4, which analyses the estimation of binary response models with fixed effects under a pseudo-panel setting. Such models are not just necessary for analysing poverty but other important issues for which there is extensive RCS data on such as unemployment, health status, or educational attainment. In the study four different pseudo-panel estimators are considered; the linear estimator considered in the previous chapter (EWALD) and three nonlinear estimators adapted from the literature on nonlinear fixed effects panel models. Results from Monte Carlo simulations and an empirical example show that pseudo-panels can consistently estimate binary response models but with stricter requirements for cohort construction than for linear models, especially if regressors are binary or inter-related. EWALD is found to be generally more robust than the nonlinear estimators when there are difficulties in meeting these requirements.

The most significant contribution of this thesis are the measures (AWAR and CAWAR) created to assess cohort construction to see how reliable pseudo-panel estimates are likely to be. Chapters 3 and 4 show the power of the measures both in a controlled Monte Carlo setting, in an applied setting using real data, and for linear and nonlinear models. We believe such measures are a much needed addition to the literature where pseudo-panel models are increasingly being used, especially in the

field of development, but with little thought given to the cohort construction process. This is particularly true as Chapter 3 has shown that many of the common methods of constructing cohorts are likely to produce heavily biased results. The chapter also highlights how such a measure can be utilised to accept or reject a cohort specification, giving the necessary thresholds required to do this. We recommended that all regressors in the model need to meet the threshold for the cohort specification to be accepted. Doing so left only four specifications from the 65 we initially started with. This indicates that passing the threshold may be quite difficult in practice, especially as we only considered AWAR in our empirical application rather than CAWAR. The latter measure, which should be applied when RCS data is used, includes cell sizes whereas AWAR does not. This means there will be a trade-off when trying to increase CAWAR to meet the threshold between increasing cell size and increasing the number of cohorts (as increasing the number of cohorts leads to greater cohort level variation). Hence it is very possible that no cohort specification exists where all the regressors meet the CAWAR cut-off. We now discuss some possibilities for addressing this.

One important extension is to assess how pseudo-panel estimators which include a measurement error correction (such as the errors-in-variables estimators suggested by Deaton, 1985; Verbeek and Nijman, 1993; Devereux, 2007a) affect the CAWAR thresholds. The corrections should reduce the cell sizes required to overcome measurement error bias thus leading to a lower threshold value than that which is required for EWALD. Furthermore, our analysis has shown threshold values may need to be larger for binary or inter-related regressors and binary response models. Additional estimators could be developed which include specific measurement error corrections for such models and/or regressors, further reducing the required thresholds.

In terms of immediate solutions to the problem of low CAWAR values, one could ease the way thresholds are applied. The ideal scenario is where all regressors have a CAWAR value above the relevant threshold and the whole regression can be used for inference. If this is not possible then it may be better to use different cohort specifications for different regressors. In this way the specification which produces the highest CAWAR value for each regressor is used to draw inference for that

regressor only (a similar method was successful when we wanted to reduce the number of cohort construction variables in Table 3.13 in Chapter 3). By selecting a specific specification for each regressor the number of construction variables required was reduced from five to three and consequently average cell sizes became at least two to three times larger. Similarly one could focus on using binding thresholds just for the main regressors of interest rather than all the other variables. One could also focus on having higher CAWAR values for regressors likely to suffer most from having low CAWAR, such as binary and inter-related regressors rather than the more continuous variables which are less biased even when CAWAR does not meet its threshold. However, if one is still struggling to obtain CAWAR values close to the required threshold than pseudo-panel estimates need to be interpreted with caution and it may be the case such models or regressors can only be estimated with genuine panel data.

The main limitation of the pseudo-panel studies are that the empirical examples are based on panel data rather than RCS data. This makes it difficult to empirically validate the Monte Carlo results for CAWAR and its recommended thresholds. The fact that the results for AWAR are consistent for both the Monte Carlo simulations and the application to panel data does somewhat mitigate this concern. However the inclusion of cell size into the CAWAR measure may mean how it operates in the Monte Carlo setting is different to an applied setting. This is because CAWAR can be increased just by having larger cell sizes rather than creating more cohort level variation. As there is generally a trade-off between these two, how the trade-off operates empirically may be different to how it does in our Monte Carlo scenario and thus this would affect the required thresholds we derived. The only way to find out is to test it empirically with RCS data. The problem with RCS data is that it will be difficult to identify the "true" coefficient estimates like we did with panel data and thus to assess how accurate pseudo-panel estimates are. One solution to this is use a large panel dataset where panel estimation can be conducted to find the "true" coefficients. Then one can drop individuals so that there are different set of individuals in different time periods and hence create RCS data. Finally pseudo-panel estimation can be conducted, the CAWAR values can be calculated, and estimates can be compared to the "true" panel coefficients. The panel data needs to be large enough that after dropping observations to create the RCS data there should still be a similar

amount of observations as usually found for normal RCS data, which is usually several thousand per survey wave.

Another limitation of the CAWAR measure is that the recommended thresholds are not very stable across models and even across regressor type. Our results suggest binary and inter-related regressors require higher thresholds than more continuous variables and the binary response model also required a higher threshold than the linear model. It is therefore possible other models we have highlighted which can be estimated by pseudo-panels may have different requirements as well. For example, models with lagged dependent variables may also require one to consider the CAWAR values of the dependent variable and not just the independent ones as we have done so far. At one extreme one may need to calibrate the required CAWAR threshold using Monte Carlo simulations for a specific model before conducting the empirical analysis because of the sensitivity of the thresholds to the type of model and variables. Further work is required to explore this issue, nevertheless the construction of this measure is still a useful first step into making pseudo-panel estimations more viable to estimate models not possible before due to data limitations.

One final avenue for future work which we believe has great potential is to explore whether one can combine data (whether it be RCS or panel) from completely different surveys by matching on cohorts. One example relevant for many countries would be matching household consumption/expenditure surveys with demographic and health surveys. If cohorts are truly representative of the households within each cohort, then as long as the surveys have data on the cohort construction variables it should be possible to create identical cohorts and merge the two datasets. Future work could explore if this works in practice where rather than the set of individuals within a cohort changing every wave, within a wave cohorts will have a different set of individuals for different variables. If such a merger is possible then this would be of great benefit to applied work where additional variables could be incorporated in a way not possible before, helping to bring together issues that previously had to be studied in relative isolation.

References

Alessie, R., Devereux, M.P. and Weber, G., 1997. Intertemporal consumption, durables and liquidity constraints: A cohort analysis. *European Economic Review*, 41, pp.37–59.

Angrist, J.D., 1991. Grouped-data estimation and testing in simple labor-supply models. *Journal of Econometrics*, 47, pp.243–266.

Antman, F. and McKenzie, D., 2007a. Earnings mobility and measurement error: A pseudo-panel approach. *Economic Development and Cultural Change*, 56(1), pp.125–161.

Antman, F. and McKenzie, D., 2007b. Poverty traps and nonlinear income dynamics with measurement error and individual heterogeneity. *Journal of Development Studies*, 43(6), pp.1057–1083.

Appleton, S., 1996. Women-headed households and household welfare: An empirical deconstruction for Uganda. *World Development*, 24(12), pp.1811-1827.

Arndt, C., Demery, L., McKay, A. and Tarp, F., 2015. Growth and poverty reduction in Tanzania. Helsinki: *WIDER Working Paper 2015/051*

Aryeetey, E. and McKay, A., 2007. Ghana: the challenge of translating sustained growth into poverty reduction. In: Besley, T. and Cord, L. eds. *Delivering on the promise of pro-poor growth: Insights and Lessons from country experiences*. Basingstoke: Palgrave Macmillan

Attanasio, O.P., Blow, L., Hamilton, R. and Leicester, A., 2009. Booms and busts: Consumption, house prices and expectations. *Economica*, 76, pp.20–50.

Bach, D., 2004. Inching towards a country without a state: prebendalism, violence and state betrayal in Nigeria. In: Chapman, C., Herbst, J. and Mills, G. eds. *Africa's big states*. Johannesburg: Witwatersrand University Press

Banks, J., Blundell, R. and Preston, I., 1994. Life-cycle expenditure allocations and the consumption costs of children. *European Economic Review*, 38, pp.1391–1410.

Barr, A. and Fafchamps, M., 2006. A client-community assessment of the NGO sector in Uganda. *The Journal of Development Studies*, 42(4), pp.611-639.

Bedi, A.S., Kimalu, P. K., Manda, D.K. and Nafula, N., 2004. The decline in primary school enrolment in Kenya. *Journal of African Economies*, 13(1), pp.1–43.

Bellemare, M.F., Masaki, T. and Pepinsky, T.B., 2015. *Lagged explanatory variables and the estimation of causal effects*. Manuscript.

Bernard, J.T., Bolduc, D. and Yameogo, N.D., 2011. A pseudo-panel data model of household electricity demand. *Resource and Energy Economics*, 33, pp.315–325.

Blundell, R., Browning, M. and Meghir, C., 1994. Consumer demand and the life-cycle allocation of household expenditures. *Review of Economic Studies*, (61), pp.57–80.

Blundell, R., Duncan, A. and Meghir, C., 1998. Estimating labour supply responses using tax reforms. *Econometrica*, 66(4), pp.827–861.

Blundell, R. and Powell, J.L., 2003. Endogeneity in nonparametric and semiparametric regression models. *Econometric Society Monographs*, 36, pp.312-357.

Blundell, R.W. and Powell, J.L., 2004. Endogeneity in semiparametric binary response models. *The Review of Economic Studies*, 71(3), pp.655-679.

Bowden, S., Chiripanhura, B. and Mosley, P., 2008. Measuring and explaining poverty in six African countries: A long-period approach. *Journal of international development*, 20(8), pp.1049-1079.

Bowden, S and Mosley, P., 2012. Four African case studies: Ghana, Uganda, Kenya and Zimbabwe. In: Mosley, P. ed. *The Politics of Poverty Reduction*. Oxford: Oxford University Press

Browning, M., Deaton, A. and Irish, M., 1985. A profitable approach to labor supply and commodity demands over the life-cycle. *Econometrica*, 53(3), pp.503–543. Available at: http://www.jstor.org/stable/1911653.

Cameron, A. C. and Trivedi, P.K., 2008. *Microeconometrics: Methods and Applications.* Cambridge: Cambridge University Press

Campbell, J.Y. and Cocco, J.F., 2007. How do house prices affect consumption? Evidence from micro data. *Journal of Monetary Economics*, 54, pp.591–621.

Chamberlain, G., 1980. Analysis of covariance with qualitative data. *Review of Economic Studies*, 47, pp.225-238

Chamberlain, G., 1982. Multivariate regression models for panel data. *Journal of econometrics*, 18(1), pp.5-46.

Chen, S. and Ravallion, M., 2010. The developing world is poorer than we thought, but no less successful in the fight against poverty. *The Quarterly Journal of Economics*, 125(4), pp.1577-1625.

Christiaensen, L.J. and Subbarao, K., 2005. Towards an understanding of household vulnerability in rural Kenya. *Journal of African Economies*, 14(4), pp.520–558.

Collado, M.D., 1998. Estimating binary choice models from cohort data. *Investigaciones Economicas*, 22(2), pp.259-76.

Cuesta, J., Ñopo, H. and Pizzolitto, G., 2011. Using pseudo-panels to measure income mobility in Latin America. *Review of Income and Wealth*, 57(2), pp.224–246.

Dang, H.A., Lanjouw, P., Luoto, J. and McKenzie, D., 2014. Using repeated cross-sections to explore movements into and out of poverty. *Journal of Development Economics*, 107, pp.112–128.
Available at: http://dx.doi.org/10.1016/j.jdeveco.2013.10.008.

Dargay, J.M., 2002. Determinants of car ownership in rural and urban areas: A pseudo-panel analysis. *Transportation Research Part E: Logistics and Transportation Review*, 38, pp.351–366.

Dargay, J.M. and Vythoulkas, P.C., 1999. Estimation of a dynamic car ownership model: A pseudo-panel approach. *Journal of Transport Economics and Policy*, 33(3), pp.287–302.

Deaton, A., 1985. Panel data from time series of cross-sections. *Journal of Econometrics*, 30, pp.109–126.

Deaton, A., 2005. Measuring poverty in a growing world (or measuring growth in a poor world). *Review of Economics and statistics*, 87(1), pp.1-19.

Deaton, A. and Paxson, C., 1994. Intertemporal choice and inequality. *Journal of Political Economy*, 102(3), pp.437–467.

Deaton, A. and Zaidi, S., 2002. Guidelines for constructing consumption aggregates for welfare analysis (Vol. 135). World Bank Publications.

Devarajan, S., 2013. Africa's statistical tragedy. *Review of Income and Wealth*, 59(S1), pp.S9-S15.

Devereux, P.J., 2007a. Improved errors-in-variables estimators for grouped data. *Journal of Business & Economic Statistics*, 25(3), pp.278–287.

Devereux, P.J., 2007b. Small-sample bias in synthetic cohort models of labour supply. *Journal of Applied Econometrics*, 22, pp.839–848.

Dollar, D. and Kraay, A., 2002. Growth is Good for the Poor. *Journal of economic growth*, 7(3), pp.195-225.

Échevin, D., 2013. Measuring vulnerability to asset-poverty in Sub-Saharan Africa. *World Development*, 46, pp.211–222. Available at: http://dx.doi.org/10.1016/j.worlddev.2013.02.001.

Fernández-Val, I., 2009. Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, 150(1), pp.71-85.

Fernandez-Villaverde, J. and Krueger, D., 2007. Consumption over the life cycle: facts from consumer expenditure survey data. *Review of Economics and Statistics*, 89(3), pp.552–565. Available at: http://www.mitpressjournals.org/doi/abs/10.1162/rest.89.3.552.

Fosu, A.K., 2015. Growth, inequality and poverty in Sub-Saharan Africa: recent progress in a global context. *Oxford Development Studies*, 43(1), pp.44-59.

Fulford, S., 2014. Returns to education in India. *World Development*, 59, pp.434–450. Available at: http://dx.doi.org/10.1016/j.worlddev.2014.02.005.

Fuller, W.A., 1975. Regression analysis for sample survey, *Sankhya: The Indian Journal of Statistics*, C37, pp.117-132

Fuller, W.A., 1981. Measurement error models (Department of Statistics, Iowa State University, Ames, IA).

Gardes, F., Duncan, G., Gaubert, P., Gurgand, M. and Starzec, C., 2005. Panel and pseudo-panel estimation of cross-sectional and time series elasticities of food

consumption: The case of American and Polish data. *Journal of Business & Economic Statistics*, 23(2), pp.242–253. Available at: http://arxiv.org/abs/0710.0559.

Gassner, K., 1998. An estimation of UK telephone access demand using pseudo-panel data. *Utilities Policy*, 7(3), pp.143–154.

Gomanee, K., Morrissey, O., Mosley, P. and Verschoor, A., 2005. Aid, government expenditure, and aggregate welfare. *World Development*, 33(3), pp.355-370.

Gómez Soler, S.C., 2016. Educational achievement at schools: Assessing the effect of the civil conflict using a pseudo-panel of schools. *International Journal of Educational Development*, 49, pp.91–106.

Glewwe, P., 1991. Investigating the Determinants of Household Welfare in the Côte d'Ivoire. *Journal of Development Economics.* 35: 307-37.

Güell, M. and Hu, L., 2006. Estimating the probability of leaving unemployment using uncompleted spells from repeated cross-section data. *Journal of Econometrics*, 133(1), pp.307–341.

Hahn, J. and Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4), pp.1295-1319.

Hancock, W.K., 1943. *Argument of empire*. Harmondsworth: Penguin Books.

Heshmati, A. and Kumbhakar, S.C., 1997. Estimation of technical efficiency in Swedish crops farms: a pseudo panel data approach. *Journal of Agricultural Economics*, 48(1), pp.22–37.

Himaz, R. and Aturupane, H., 2016. Returns to education in Sri Lanka: A pseudo-panel approach. *Education Economics*, 24(3), pp.300–311.

Iliffe, J., 1969. *Tanganyika under German rule 1905-1912*. Cambridge: Cambridge University Press.

Imai, K.S., Annim, S.K., Kulkarni, V.S. and Gaiha, R., 2014. Women's empowerment and prevalence of stunted and underweight children in rural India. *World Development*, 62, pp.88–105.
Available at: http://dx.doi.org/10.1016/j.worlddev.2014.05.001.

Jæger, M.M., 2013. The effect of macroeconomic and social conditions on the demand for redistribution: A pseudo panel approach. *Journal of European Social Policy*, 23(2), pp.149-163.

Jerven, M., 2013. Poor numbers: how we are misled by African development statistics and what to do about it. Ithaca NY: Cornell University Press.

Jerven, M. and Johnston, D., 2015. Statistical tragedy in Africa? Evaluating the data base for African economic development. *The Journal of Development Studies*, 51(2), pp.111-115.

Jiang, S.S. and Dunn, L.F., 2013. New evidence on credit card borrowing and repayment patterns. *Economic Inquiry*, 51(1), pp.394–407.

Kopiński, D., Polus, A. and Tycholiz, W., 2013. Resource curse or resource disease? Oil in Ghana. *African Affairs*, 112(449), pp.583-601.

Kosack, S., 2012. The education of nations: How the political organization of the poor, not democracy, led governments to invest in mass education. Oxford: Oxford University Press

Leibbrandt, M., Woolard, I., Finn, A. and Argent, J., 2010. Trends in South African income distribution and poverty since the fall of apartheid. *OECD, Social, Employment and Migration Working Papers 101*.

Lonsdale, J., 2012. Ethnic patriotism and markets in African history. In: Hino, H., Lonsdale, J., Ranis, G. and Stewart, F. eds. *Ethnic Diversity and Economic Instability in Africa*. Cambridge: Cambridge University Press, pp.19-55.

Martins, P., 2013. Growth, employment and poverty in Africa: tales of lions and cheetahs. *Background paper prepared for the World Development Report* 2013

McKay, A., 2013. Growth and poverty reduction in Africa in the last two decades: Evidence from an AERC growth-poverty project and beyond. *Journal of African Economies*, 22(suppl 1), pp.i49-i76.

Meng, Y., Brennan, A., Purshouse, R., Hill-McManus, D., Angus, C., Holmes, J. and Meier, P.S., 2014. Estimation of own and cross price elasticities of alcohol demand in the UK - A pseudo-panel approach using the Living Costs and Food Survey 2001-2009. *Journal of Health Economics*, 34, pp.96–103. Available at: http://dx.doi.org/10.1016/j.jhealeco.2013.12.006.

Mkandawire, T., 2010. On tax efforts and colonial heritage in Africa. *The Journal of Development Studies*, 46(10), pp.1647-1669.

Moffitt, R., 1993. Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics*, 59, pp.99–123.

Mogues, T., Morris, M., Freinkman, L., Adubi, A. and Ehui, S.K., 2008. Agricultural public spending in Nigeria. *IFPRI Discussion Paper 00789*.

Mosley, P., 2012. *The politics of poverty reduction*. Oxford: Oxford University Press.

Mosley, P., 2013. Two Africas?: Why Africa's “Growth Miracle” is Barely Reducing Poverty. *Brooks World Poverty Institute Working Paper 191*.

Mosley, P., 2015. Fiscal composition and aid effectiveness: a political economy model. *World development*, 69, pp.106-115.

Mundlak, Y., 1978. On the pooling of time series and cross section data. *Econometrica*, 46, pp.69-85.

Myint, H., 1976. *The economics of the developing countries*. London: Hutchinson University Library.

Neyman, J. and Scott, E.L., 1948. Consistent estimates based on partially consistent observations. *Econometrica*, 16(1), pp.1-32.

Nicita, A., 2009. The price effect of tariff liberalization: Measuring the impact on household welfare. *Journal of Development Economics*, 89, pp.19–27. Available at: http://dx.doi.org/10.1016/j.jdeveco.2008.06.009.

Nsowah-Nuamah, N., Teal, F. and Awoonor-Williams, M., 2010. *Jobs, Skills and Incomes in Ghana: How was poverty halved?* (No. CSAE WPS/2010-01).

Ogut, B., 2012. Ethnic patriotism and markets in African history. In: Hino, H., Lonsdale, J., Ranis, G. and Stewart, F. eds. *Ethnic Diversity and Economic Instability in Africa*. Cambridge: Cambridge University Press, pp.91-126.

Okidi, J.A., Ssewanyana, S., Bategeka, L. and Muhumuza, F., 2007. Uganda's experience with operationalising pro-poor growth, 1992 to 2003. In: Besley, T. and Cord, L. eds. *Delivering on the promise of pro-poor growth: Insights and Lessons from country experiences*. Basingstoke: Palgrave Macmillan

Olinto, P. and Uematsu, H., 2013. The state of the poor: where are the poor and where are they poorest?. Washington DC: World Bank, Poverty Reduction and Equity Department, www.worldbank.org.content/dam/worldbank/document/state_of_the_poor_paper_April17.pdf

Papke, L.E. and Wooldridge, J.M., 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, 145(1), pp.121-133.

Paul, C.J.M. and Nehring, R., 2005. Product diversification, production systems, and economic performance in U.S. agricultural production. *Journal of Econometrics*, 126, pp.525–548.

Pinkovskiy, M. and Sala-i-Martin, X., 2010. African Poverty Is Falling... Much Faster Than You Think! *NBER Working Paper*, *15775*.

Porter, G., 2003. NGOs and poverty reduction in a globalizing world: perspectives from Ghana. *Progress in Development Studies*, 3(2), pp.131-145.

Poteete, A.R., 2009. Is development path dependent or political? A reinterpretation of mineral-dependent development in Botswana. *The Journal of Development Studies*, 45(4), pp.544-571.

Propper, C., Rees, H. and Green, K., 2001. The Demand for Private Medical Insurance in the UK : A Cohort Analysis. *The Economic Journal*, 111(471), pp.180–200.

Ravallion, M., 2000. Should poverty measures be anchored to the national accounts? *Economic and Political Weekly*, pp.3245-3252.

Reed, W.R., 2015. On the practice of lagging variables to avoid simultaneity. *Oxford Bulletin of Economics and Statistics*, 77(6), pp.897-905.

Rivers, D. and Vuong, Q.H., 1988. Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, 39(3), pp.347-366.

Sprietsma, M., 2012. Computers as pedagogical tools in Brazil: a pseudo-panel analysis. *Education Economics*, 20(1), pp.19–32.

Stock, J.H. and Yogo, M., 2005. Testing for Weak Instruments in IV Regression. In: Andrews, D.W.K. and Stock, J.H. eds. *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas Rothenberg*. Cambridge University Press, pp.80–108.

Terza, J.V., Basu, A. and Rathouz, P.J., 2008. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27(3), pp.531-543.

United Nations Development Programme (UNDP), 2013. Assessing Progress in Africa toward the Millennium Development Goals. *MDG Report 2013*, United Nations Development Programme and Economic Commission for Africa.

Verbeek, M. and Nijman, T., 1992. Can cohort data be treated as genuine panel data? *Empirical Economics*, 17, pp.9–23.

Verbeek, M. and Nijman, T., 1993. Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections. *Journal of Econometrics*, 59, pp.125–136.

Verbeek, M., 2008. Pseudo Panels and Repeated Cross Sections. In: Matyas, L. and Sevestre, P. eds. *The Econometrics of Panel Data: Fundamental and Recent Development in Theory and Practice Third Edition*. Berlin: Springer, pp.369–383.

Warunsiri, S. and McNown, R., 2010. The returns to education in Thailand: A pseudo-panel approach. *World Development*, 38(11), pp.1616–1625. Available at: http://dx.doi.org/10.1016/j.worlddev.2010.03.002.

Wooldridge, J. M., 2010a. Econometric Analysis of Cross Section and Panel Data. 2nd ed. The MIT Press.

Wooldridge, J. M., 2010b. Correlated random effects models with unbalanced panels. Manuscript.

Wooldridge, J.M., 2014. Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics,* 182(1), pp.226-234.

World Bank, 2011. *Tackling poverty in Northern Ghana*. Washington, DC: World Bank.

World Bank, 2013a. The state of the poor: where are the poor, where is extreme poverty harder to end, and what is the current profile of the world's poor? Economic premise; no. 125. Washington DC; World Bank.

World Bank, 2013. Nigeria: Poverty in the midst of plenty: the challenge of growth with inclusion. Washington, DC: World Bank.

World Bank, 2015. *Tanzania Mainland Poverty Assessment*. Washington, DC: World Bank.