Household formation in post-apartheid South Africa, 1995-2011: Measurement and Trends

Amy Julia Thornton

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the School of Economics

UNIVERSITY OF CAPE TOWN

Supervisor: Prof. Martin Wittenberg

March 2021

Declaration of Own Work

I, Amy Julia Thornton, declare that this is my own original work in concept and execution. I have received no assistance besides the normal guidance from my supervisor. No part or version of this thesis has been submitted at any point for the award of a degree at this, or any other, university.

Signed:

Amy Julia Thornton
1 March 2021
Cape Town

To Mom, Dad, and, Wayne

Abstract

Who is in your household, what they have to share with you, and who does what, has profound implications for your welfare and well-being. Child welfare and progress through school; continued female disadvantage in the labour market and gender-based violence; livelihood strategies and physical health - these are just a few social outcomes that are fundamentally structured by the household. Change in most outcomes, therefore, cannot be abstracted from change in households.

In South Africa, households have been changing in systematic ways. Between the first census of the post-apartheid period in 1996, and the latest in 2011, the number of households increased by 62% from 9 million to 14.5 million. By contrast, the population only grew by 28%, so that the average household shrank by almost a whole person. A notable aspect of this change is an extensive increase in the rate at which people live alone: the number of households that were single-person increased by 160% over the same time period. In other words, over time the South African population has been spread more thinly over fewer households, with direct implications for a sweeping set of social outcomes of importance to the post-apartheid project.

A large literature on the topic of household composition exists for South Africa, and many researchers have noted the decline in average household size. Relatively fewer studies, however, aim to uncover the process by which this is happening, and those that do, mainly rely on highly localised data. Indeed, the question of why South Africans would form more and smaller households is particularly provoking given that economic circumstances have arguably been challenging in the post-apartheid period, characterised by high and persistent levels of unemployment and extreme wage inequality. As such, this thesis set out to investigate the process of household formation in the post-apartheid period, and how this aligned with known drivers of household formation in the economic literature, being aging, employment, and marriage.

To do this, we construct a harmonised series of nationally-representative cross-sectional household surveys collected (almost) annually by Statistics South Africa, covering the period 1995-2011. Our first contribution is to make progress towards overcoming serious data quality issues in these surveys that undermine not only our study, but most studies seeking to relate people to households in data. We then turn towards analysing household formation, which raises the question of how to measure this process. We approach household formation from two angles: firstly, the aggregate household count by studying household heads, and secondly, the process of leaving the parental household by studying young adults.

By tracing trends in household headship, we are able to describe which groups have become more or less likely to form households, and how this changed over the period between 1995 and 2011. Thanks to our attention to data quality, we are able to use the surveys to do the same for single-person households, a sub-group of special interest, in a way that is reliable for the first time. Our third contribution is to describe macro level trends in leaving the parental household, also for the first time for South Africa. Our key finding is that household formation in post-apartheid South Africa has been profoundly impacted by the steady decline of marriage over the same period. As such, household formation patterns are highly gendered, and modulated by men and women's differential access to labour market and grant income, over the life-cycle.

Acknowledgements

Firstly, I am grateful to the David & Elaine Potter Foundation and the University of Cape Town for their generous funding of my doctoral degree. Without this funding, I would not have been able to embark on this process at all, and certainly not carve out time to work exclusively on my research when the time came for it.

To my supervisor, Prof. Martin Wittenberg, thank you for your continued guidance and support, even in sometimes difficult circumstances for both of us, and not least, in the middle of a global pandemic. It was important to me from the beginning to produce something I was proud of and I knew you would be the best person to help me achieve this. Thank you for reading, correcting, commenting, meeting, discussing, and more, over and over again. Thank you for your time.

Thanks are also due to my colleagues at the Development Policy Research Unit where I worked part-time throughout my PhD. Thank you to Prof. Haroon Bhorat and Dr. Morne Oosthuizen for supporting me taking the time to embark on the PhD and easing the process of juggling work and research at the same time. Thanks also to my other friends at the DPRU who offered advice and moral support on the basis of their own PhD experiences.

I would also like to specially mention Jacqueline Mosomi, a friend who 'went before' on the PhD journey, and who was incredibly generous in sharing her experience, providing guidance, and helping me know what to expect. Thank you for long chats, answering questions, and listening. You have been invaluable.

Heartfelt thanks also to my parents, John and Justine Thornton, for your unwavering support and belief in me from the very beginning. Thank you for the opportunities you provided for me, and for imparting the value of education and hard work, set by your own good example. Thank you also for creating an environment that helped me achieve this, especially during the pandemic. Writing this thesis would not have been possible without you.

Lastly, thank you to my husband, Wayne Eldridge. For the past three years, this thesis has been as much a part of your life, as it has been a part of mine. You have been a phenomenal source of support, providing me with the time and space I needed to work on this project, picking up my slack in our everyday lives without complaint. Thank you for cups of coffee, doing dishes, buying groceries, checking I'd eaten, more coffee, and a hundred other small, but hugely significant, acts of love and support. Without you, this process would not only have been indescribably harder, but I suspect I would have come out the other end, as quite a different - substantially less resilient - person. Thank you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Who you know matters

Who you know and how much they have matters for almost every outcome social scientists are interested in studying. The size and structure of your social network - being family, friends, classmates, colleagues, etc. - affects outcomes ranging from education to work to health and more. Your social connections affect whether you attend school and for how long (Hofferth et al., 1998); whether you study further (Sandefur et al., 2006) and what you choose to study if you do (Andres et al., 1999). Who you know affects your future earning potential (Lam and Schoeni, 1993; Gouskova et al., 2010), how easily you will find a job (Schöer et al., 2014), or access capital to start a business (Shih, 2004; Musinguzi, 2016). Your social network determines who you meet and potentially marry (Lichter and Qian, 2019; Schwartz, 2013); where you live (Dawkins, 2006) and your ability to migrate for work (Awumbila et al., 2017). Who you know affects your values (Helman and Ratele, 2016); your religious beliefs (Hoge et al., 1982); your politics (Aggeborn and Nyman, 2020); and ultimately your physical and mental health and well-being (Holt-Lunstad et al., 2010; Pietromonaco and Collins, 2017). This is because social networks are the infrastructure by which people share information and resources (Blumenstock et al., 2019), meaning who is in your network and how many resources they have to share with you has profound implications for your life outcomes. By studying the household, the clearest lens we have in our data on social connections in South Africa, this thesis is documenting seventeen-years-worth of arguably the most critical context of the lives of South Africans with ramifications for a broad spectrum of their social outcomes.

The research in this thesis reveals that who people live with has changed dramatically over the post-apartheid period. People's households are getting smaller, not only because they are having fewer children but also because of highly gendered changes in household formation that appear to be driven to some extent by the decline in marriage. As people have become less and less likely to ever get married (Garenne, 2016), so the rate at which never-married men and women are forming new households has increased. There has been an astronomic rise in the rate at which men live alone in South Africa accompanied by

an increase in female-headed households that account for the lion's share of childcare and increasingly include extended family members. We find that in a context of high youth unemployment and a social welfare system that ignores them, young adults and young women especially continue to rely on the parental household as a private security net. These gendered patterns have resulted in more men and women living in households comprised of adults who are male or female only, in part because so many more people are living alone. Who does what and decides what in the household is one of the most important ways in which livelihoods and gender relations are structured. Understanding changing gender relations could not be more important in a country where gender-based violence has been described as an 'epidemic' by media and researchers alike (Abrahams et al., 2009; Karim and Baxter, 2016) and where most children grow up without a co-resident father (Hall and Posel, 2019). As such, documenting the evolution of people's social context is of primary importance anywhere, but perhaps especially so in South Africa where the social fabric has so recently undergone change of the most fundamental order.

1.2 Setting the scene

South Africa underwent landmark political change in 1994 with its first democratic election. Prior to this, South Africa had been under apartheid rule during which a white minority government subjected the country to decades of racist social engineering. South Africans were classified into four racial categories which determined how the State would treat them: Black African, Coloured, Asian/Indian, and white. The apartheid regime systematically impoverished the Black[1] (African, Coloured and Asian/Indian) population for white enrichment by exploiting Black labour, restricting Black freedom of movement, and forcing Black populations to live further away from economic opportunities (Mabin, 2003; Lemon, 1991). After the democratic election, a society based on a racist hierarchy transitioned to one that valued the freedom and human rights of all South Africans. A constitution was set up reflecting this change which was amongst the most liberal and far-reaching in the world in terms of the entitlements and freedoms it afforded South Africans (Comaroff and Comaroff, 2003).

However, when the country embarked on democracy in 1994, the effects of apartheid social engineering were already deeply entrenched. Many of the freedoms afforded to people on paper in the constitution were restricted by economic realities. High up on the agenda for the incoming post-apartheid government therefore was combating the three ills of high open unemployment, widespread poverty, and a level of income inequality that was amongst the worst in the world (Wittenberg, 2017). Black unemployment had shot up in the homelands in the 1970s (Nattrass and Seekings, 2010) and was exacerbated when the mining industry, a centrepiece of the apartheid economy, underwent a series of crises in the 1980s so that by the early 1990s it was shedding jobs at an extraordinary rate (Wilson, 2001). By 1994, 21% of Black African men were unemployed (Casale et al., forthcoming). This had direct consequences for welfare since apartheid labour market engineering and land dispossession meant (and still means) that a job is often the difference between falling into poverty or not (Woolard and Klasen, 2005). More than half the population counted as living in poverty in 1995 (Bhorat and van der Westhuizen, 2012; DPME,

[1]We use the term 'Black' to apply to groups of the population classified as 'African', 'Coloured' (people of mixed-race heritage from the Cape and also associated with a distinct cultural identity); and 'Indian/Asian'. These are apartheid-era classifications which Statistics South Africa still uses to collect demographic data owing to the continued importance of understanding, quantifying and monitoring the legacy of apartheid and the disadvantages it entrenches in the post-apartheid era. The fourth population group on which Statistics South Africa collect data is 'white'.

2014). Concentrated wage inequality also put limits on how far access to labour earnings could take a family. The wage Gini coefficient was just below 0.5 in 1994 (Wittenberg, 2017), and most wealth, skills, and capital was in the hands of whites.

To address these problems, the new government's strategy had a wage and non-wage prong. The non-wage prong consisted of firstly a generous programme of social expenditure in the form of cash grants targeted mainly at the aged, the infirm, and children, but also the free provision of housing and access to a base level of electricity and water. Between 1996 and 2013, the total number of social grant recipients grew from 2.4 million to just under 16 million (DPME, 2014), and by 2009 the government claimed to have built 2.6 million free houses (President Motlanthe, 2009). The wage prong was integrating Black people more equitably and effectively into the core of the labour market from which they had been systematically excluded for so long. This was the much bigger challenge since it required confronting racially imbalanced distributions of ownership, education and skills that had accumulated over generations. South African society had been fundamentally altered in 1994 and the new goal of prosperity for all was not without serious incumbent challenges.

1.3 A post-apartheid research agenda

At the forefront of the post-apartheid research agenda then was the collection, measurement and monitoring of a host of human development and labour market indicators to track the progress of the post-apartheid project. Information needed to be accumulated on all of these topics in order to assess the scope of the challenges being faced, as well as the best routes by which to start tackling them in order to uplift the welfare of the population. Yet, when only 9 million out of 40 million South Africans were employed in 1994 (Casale et al., forthcoming), who got a portion of those 9 million paychecks depended largely on who you knew and how you were connected. In South Africa, welfare is often measured by comparing per capita household income to a poverty line, meaning we really measure household welfare and make an individual approximation. A person's welfare status will then depend directly on whether the income they are bringing in has to be spread across few or many people, and if any of those people are unemployed. This was particularly because, in the absence of State-funded unemployment support, the large unemployed population relied on households as private safety security nets (and continue to do so, today) (Klasen and Woolard, 2009). Similarly, connections anyone might have outside the household might serve to bring remittances into the household or represent ties that draw some of their income outside of it.

Questions about welfare, therefore, cannot be abstracted from the context in which someone lives; this is true about the vast majority of other social outcomes too, and evidence that this is the case exists for South Africa. For example, living in the same household as a grandmother receiving a pension in South Africa has been shown to improve child's height-for-age, especially girls (Duflo, 2000). In the country's anaemic labour market, social networks are the most important way in which people source a job (Schöer et al., 2014), and people who live with other employed people are more likely to be employed themselves (Pirouz, 2005). Social connections also proved invaluable when South Africa came under the grip of the HIV/AIDS epidemic. The demographic hit hardest by AIDS was young Black African women, meaning grandmothers become the safety net for millions of AIDS orphans (Nyasani et al., 2009), as well

as, performing a palliative care role for the many sick people who returned home to be cared for (Clark et al., 2007). Therefore, a key part of studying change over the course of the post-apartheid period in most social science outcomes is understanding how people's social connections and context have changed.

1.4 Measuring changing social contexts

The main way we can do this is by using the household survey data collected since the end of apartheid. Part of the drive to track the progress of the post-apartheid project and to enable evidence-based policy-making was the collection of household survey data by the national statistics bureau. Every year since 1994, cross-sectional household survey data representative of the whole population at once (including integrating those in former independent homelands) is collected. These surveys are designed to collect data on outcomes that will inform government's economic planning and capture information ranging from labour market to socio-demographic outcomes. The most important way in which social context or social connections are captured in these data is the household. One of the first questions respondents are asked is to identify a household head and then all other household members are defined by their relationship to this person: spouses, children, siblings, relatives, or other types of members. The way in which these household connections are captured has remained largely unchanged over time, providing us with a snapshot of household life for (almost) every year of the post-apartheid period. We can use this lens to describe and document the shifting social connections of South Africans.

This lens is by no means without limitations. There are shortcomings of how well the household is being measured. The criteria for being a household member is having slept on average four nights a week in that household in the four weeks prior to enumeration. Given well-established ideas about stretched households in South Africa, where an urban migrant lives away from a rural homestead but retains his membership of that homestead (Spiegel et al., 1996), this definition could be described as strict. Social connections other than the household are also being missed: friends, classmates, work colleagues, and, in particular, non-co-resident family. Family ties are only observable in the data to the extent that they exist within the same household. If anything, this makes co-residency as the only social connection consistently and invariantly captured in our data on a long-term basis even more important to understand.

This thesis then uses this series of snapshots captured by household survey data as a lens for understanding people's changing social context over the course of the post-apartheid period. One of the most important changes to document and understand has been rapid household proliferation. Between the first census of the post-apartheid period in 1996 and the most recent census in 2011, the number of households in the country increased from 9 million to 14.5 million, a change of 62%.[2] By contrast the population only grew by 28%, from 41.3 million to 51.6 million. An immediate consequence of this is that households have become smaller: average household size has shrunk from 4.6 to 3.6 people, a drop of 22%. At the extreme, the number of people living alone mushroomed from 1.48 million in census 1996 to 3.86 million by census 2011. This represents important change in the size, but also the structure, of people's co-residency. The same young man who ten years earlier might have lived with a spouse, is much more likely to be unmarried and living alone if he lived in 2011. The woman who might have been living in his household as a spouse in 1995, is more likely to head her own household if she lived in 2011,

[2] All census statistics are own calculations using the census ten per cent samples.

made possible by her better access to a job or grant income. What might have been a single household in 1995 is thus split, with implications for who is bringing what types of income into the household; who is sharing income with whom; who can be relied upon to care for whom; and, ultimately, how men and women and children all relate to each other.

This process is mechanically driven by the household formation process - over time more people are splitting off from existing households and forming new ones. Most literature on household change in South Africa and around the world has approached the topic from the angle of household composition: what types of members live in the household and how has this changed over time? This is a related but different angle to household formation. We are interested in the process by which South Africans were spread more thinly over more households by the end of our period of study in 2011. This expansion of the household count is, we feel, an important (and less-studied) change that cascades, bends back and interacts with household composition. Of course, we do not need an ever-increasing number of households to see changes in composition; people can reorganise themselves across the same number of households, re-shaping the distribution of household size, and altering average household size in the process. But, around the world and in South Africa, the household count *is* increasing, which means more people are splitting away from existing households (what might be behind this process?) and this process unequivocally modulates the 'options', so to speak, for the shrinking remainder that need to be distributed across these households (how are these 'options' evolving?).

Understanding the drivers and consequences of household formation is important not only because household change is related to the sweeping set of socio-economic outcomes discussed above, but also because a move to smaller households doesn't necessarily gel with what we we know so far about South Africa. Firstly, it is not immediately obvious that more and smaller households will be welfare-increasing. Smaller households undermine household economies of scale making it comparatively more expensive for household members to access household public goods (Fafchamps and Quisumbing, 2007). In one part of the country, the State's ability to reduce the backlog of households not connected to the electricity grid was undermined because people were forming new households so quickly (Harris et al., 2017). Secondly, the main body of literature on household formation in South Africa informs us that employment is a key reason people form households (Keller, 2004; Ebrahim et al., 2013). But we also know that unemployment is excessive, has worsened over time, and that the unemployed tend to remain at home, drawing on the household as a private safety security net without State-sponsored unemployment support (Klasen and Woolard, 2009). It is not obvious from these findings either that household formation in South Africa would proliferate when unemployment is as high as it is.

At the same time, other pertinent changes have occurred over the period, such as the expansion of the State's social grant spending, the provision of free housing and services, and a series of expansions in women's rights in the 90s which could be enabling factors for more household formation. South Africa is also not an outlier: people have been forming households more quickly than the population has grown all around the world for decades (even centuries) and this has been the topic of an extensive literature in the fields of family sociology, demography, and economics. This literature has much to offer on the subject of household change, including identifying patterns of delayed marriage and the ideational change of the Second Demographic Transition valuing the Individual over the Family as primary drivers. We trace these developments along with providing more context about the South African case and the rich literature on households in South Africa in the literature review in the next chapter. Ultimately, in this

thesis we are interested in how we can use household survey data to learn about how household formation has changed in post-apartheid South Africa - can we measure it properly; who has become more or less likely to form new households; and, how does this connect to what we already know about the labour market, marriage market and social spending?

1.5 Contributing to understanding household formation

Throughout this thesis, we use the same stacked series of cross-sectional household surveys from Statistics South Africa (StatsSA) to study different aspects of household formation for (almost) every year between 1995 and 2011. The series comprise the October Household Surveys (1995-1999) and the General Household Surveys (2002-2011). These surveys are not designed to be used as a time-series and the first part of the thesis details some of the work that had to be put into constructing and harmonising the surveys for this purpose. In this thesis, household formation is studied using two outcomes: household headship (with those heading single-person households as a sub-group of special interest) and leaving the parental household. The thesis is then divided into four parts. Part I comprises two chapters on data quality; Part II is made up of two chapters on household headship and living alone; in Part III, there is one chapter on home-leaving; and Part IV includes one chapter considering the implications of our main findings, and the concluding chapter.

Part I details data quality issues in our stacked series of StatsSA data that complicate most research questions, but questions about households and household formation, in particular. Accurate measurement of the phenomenon we want to investigate is a prerequisite before we can embark on more substantive questions. It is important that the household change we see in our household survey data reflect that found in census data - the most reliable source for measuring household change. As it turns out, a range of data quality issues undermine the reliability of the trends we can extract from the surveys, including for main variables of interest like household counts, and cast doubt on subsequent analysis. We therefore spend the first part of this thesis addressing these issues as best we can, making the extraction of accurate trends themselves a first order output of this research, and shoring up the foundation on which later analysis is built.

The first of these data quality issues concerns the process by which StatsSA calibrated the survey weights. Due to time and capacity constraints, StatsSA split the weights for the same sample of people and households into two separate weights - one for people in the person file and one for households in the household file - which are calibrated on different input information and are never equal to each other. This creates a series of statistical and conceptual problems, including that one can reach different estimates of the total number of households for the same year depending on whether one uses the person or household weight. This introduces a great deal of uncertainty into answering a simple first-order question such as, 'how many households are there in South Africa in a given year?'. Answering this type of question is of course germane to extracting trends. The task of Chapter 3 then, is to resolve this problem by constructing a new series of integrated weights that combines both person and household information in a single weight and coherently links people to the households in which they live.

Although solving the problem of the 'dual-weight' system takes us some part down the road to extracting reliable trends, more work remains. In the process of constructing and testing out our new series

of weights, it became clear that there were other problems with the weights not specifically related to the separation of the weights into person and household weights. Instead, these issues related to under-sampling and benchmark sources. For example, single-person households - a constituency of particular interest for this thesis - have been chronically undersampled in our data. In other cases, the effect of using the 2007 Community Survey as a weighting benchmark, even though it is not a census, visibly distorts the trends. And, the lack of any benchmarking on household counts at all in the October Household Survey period results in irregular trends. Chapter 4 then augments the work of Chapter 3 by updating the weight series to account for these other weaknesses. The final outcome after both of these chapters are complete is (a) a reliable and coherent series of weights pegged to both person and household benchmarks; (b) an accurate series of household counts; and (c) in particular, an accurate series of counts of single-person households.

Having set up up data, the work in Part II delves into the substantive project of describing trends and sources of change in household headship and single-person living between 1995 and 2011. We focus on household heads because there is one head per household in our data meaning the number of heads tallies with the number of households. This provides us with a useful lens to understand country-level change by asking who headed the 9 million households in 1995, and how were they different or similar to those who headed in the 14.5 million households in 2011? Single-person households are also a group of special interest because they are the household type that has grown the fastest over the period, and contribute most quickly to household proliferation. In Chapter 5 we track how these changes relate to trends in the labour market, marriage market and social expenditure and find change in household formation is gendered. Both men and women contributed to more household formation at different times in the post-apartheid period and in different ways. Female headship boomed in the 1990s, coinciding with better female access to the labour market, the expansion of grant income access, and, a series of laws augmenting women's rights; thereafter, female headship plateaued in the decade of the 2000s. Men followed a slow-boil pattern that was level in the 90s; but then, whilst other groups of men reigned in their household formation, Black African male household headship started to gain momentum in the 2000s. Black African men are the main group behind the expansion of single-person households, no doubt related to South Africa's long history of migrant labour, but likely also linked to change in marital patterns. A key connection that starts taking shape in Chapter 5 is between the steady decline in union formation and the rise of the never-married head.

Delayed marriage has been linked to more household formation, and especially single-person household formation, in the developed world (Craigie et al., 2018; Raymo, 2015). However, in the South African econometric literature on household composition or formation, marriage has been omitted since researchers have been occupied with what seems like the more pressing question of where South Africa's large unemployed population are living (Klasen and Woolard, 2009). It is much easier to rationalise marriage being a deciding factor for household formation in the developed world where almost everyone has access to a job; less so in South Africa where a job is far from guaranteed. Chapter 6 then builds and develops the ideas that emerged from the trend analysis in Chapter 5 taking a more econometric approach and aiming in particular to try to separate out the strands of marriage and work. We set up a model of household headship and living alone, and then decompose these models over time to descriptively pick out the most important sources of change and whether the change was owed to compositional or behavioural factors. The most robust and consistent result emerging from this analysis is an acceleration in the rate

at which South Africa's growing share of never-married people are forming households. As people have become less and less likely to form a union, the rate at which never-married people head households and live alone has increased.

Leaving the parental home is a natural focal point in the literature on household formation because the 'demographically dense' (Rindfuss, 1991) period of young adulthood often intersects with getting married or getting a job for the first time (Ermisch, 1999; Goldscheider and DaVanzo, 1989). Studying young adults and their home-leaving behaviour can then better embed our findings so far about household formation in general in the context of the life-cycle. In fact, not much is known about home-leaving in South Africa since the topic has not received much scholarly attention beyond some original work by Charles Simkins (Simkins, 2017*a*,*b*), strongly motivating an empirical investigation of the topic. We turn to the question of to what extent home-leaving patterns might be contributing to more household formation in Chapter 7 in Part III. We find young women to be highly reliant on marriage as a route out of the home, making it unlikely they are contributing to more household formation. Instead it is their older counterparts with better access to the labour market who have been the driver of more female household formation. Young men, on the other hand, do appear to be contributing to more household formation, both because there has been a gentle incline in the rate at which young Black African men detach from the parental household and because they are more likely to either head their own household or live alone when they do.

In Part IV, we reflect on the implications of these findings in Chapter 8. We use our reweighted data to chart how household composition has evolved and find that there has been growth at the 'poles' of complexity: more people are living alone *and* more people are living with extended kin. The former is largely the case for men, while women are mainly forming households that include children and other female extended kin. As such, it appears that household formation patterns are reproducing the double burden of work and childcare for women, whilst men increasingly live alone with implications for their own economic welfare and physical and mental well-being. A central debate about South African household change - which we detail more in the next chapter - is to what extent South African households are becoming more nuclear. Implicit in this debate, are questions about value change and gravitation towards Western ideals. Our finding that the household formation *behaviour* of never-married people changed so extensively brings this type of question to the fore. Taking household change patterns we've uncovered as a whole, so clearly influenced by marital decline, makes us cautious to conclude the growth of single-person households amongst Black African men means they are now subscribing to nuclear family patterns. We contrast this values-explanation with a 'pattern of disadvantage' explanation incorporating the legacy effects of apartheid policies, oscillating labour migration, and continued economic precarity in the post-apartheid period. Whilst both explanations no doubt play their part, we speculate the latter is playing the dominant role in undermining marital prospects and never-married people are adjusting their household formation behaviour in turn, a mechanism for which some evidence already exists for African American women in the United States (Craigie et al., 2018). We close in Chapter 9 by reviewing what we have learnt; clarifying the contribution made by each part of this thesis to both the South African and international literature; reflecting on the limitations of our analysis; and, raising more questions for future research.

1.6 Who is forming households (and who you live with) is changing

Just like most populations around the world, South Africans are forming households more quickly than the population is growing and the result is that average household size has been shrinking. These changes are owed in part to falling fertility, but also to real change in the rate at which South Africans are forming households: households include fewer children, but also fewer adults. These changes can be tracked using household survey data as a series of successive snapshots of South African co-residency. Over time, these surveys reveal change in who is most likely to head households and crucially, who South Africans are most likely live with (or without). A man and women aged around 30-35 who might have lived together as a married couple with a child in 1995, are much less likely to do so if they were 30-35 in 2011. In 2011, both are more likely to have never married and instead he will probably be living alone and she might be heading her own household with her child and extended female kin, or else living in a similar household. Who does what in the 1995 versus 2011 household is quite different in terms of who the breadwinner is, the day-to-day household management, and household decision-making with consequences for the welfare of individual household members, their children, and the norms children grow up with and carry into their adult lives. Change in most social outcomes of interest for the post-apartheid project cannot be de-linked from the changes in household formation described in this thesis. That households are getting smaller; that men are increasingly living alone; that marriage is less and less a springboard for household formation, are all factors with cascading implications for livelihood strategies, physical health, mental health, education attainment, gender relations, child well-being, and more.

Chapter 2

Literature review

The academic literature on households is vast since it cuts across disciplines so fundamentally. For this review, our goal is to provide the background for the next chapters by positioning our project in the international literature and providing the required context for the South African case. We begin with a brief exposition of the economic theory. This is the best place to start because economic thinking around households provides us with the most powerful tools for understanding household change. Following this we provide an overview of the global decline in household size and trace thought on this topic from various disciplines, including family sociology, demography, and economics. We then turn to the South African case and begin by briefly providing the historical background that serves as such critical context for understanding contemporary household change in the country. We end by reviewing the rich literature on households in South Africa, also from a range of disciplines, touching specifically on the role of gender, and describing how this thesis contributes to this and the international literature. The themes and background discussed in this literature review thread throughout the following chapters.

2.1 Theory and international literature

2.1.1 Households in economic theory

Economic theory provides powerful tools for understanding household behaviour. The household, the traditionally female sphere, has only been seriously considered by microeconomists since Gary Becker and Jacob Mincer pioneered the 'New Home Economics' in the 1960s (Heckman, 2015; Grossbard, 2006; Pollak, 1985). Prior to this, the household had not been within the formal ambit of economists, even if some (e.g. Kuznets) had acknowledged its omission was probably important (Heckman, 2015); and when the household was included, it was modelled as a 'black box' (Pollak, 1985). The contribution of Becker and Mincer was to apply the tools of microeconomics, specifically consumer price theory, to explain household and other demographic behaviour, such as fertility (Becker, 1960). One of the key

breakthroughs enabling the development of these ideas was the conceptualisation of time as a resource with wages as it's price, as exemplified in Mincer (1962)'s work on the labour supply of married women and Becker (1965) on the allocation of time.

In his classic formulation of the household, Becker (1981) treats married couples as a firm who employ household and labour market production to maximise a household utility function. This yields different solutions for men and women since, although male and female labour is modelled as substitutes, women's biological advantage in reproduction results in a 'tipping' solution where men specialise in labour market production and women in the household. The key update of Becker (1981)'s ideas has been the recognition that households are made up of individuals and therefore to start with individual, as opposed to household utility. Ermisch (1981), for example, models the household as a technology that transforms market goods, non-market time, and household composition into individual utility. The utility of household members is impacted by household size in three ways: the pooling of expenditure to buy goods that are locally public, but congestible, e.g. housing; the pooling of time to produce household services, e.g. food preparation; and, preferences for privacy.

Other theorists have drawn from game theory and instead of defining the household as a firm or technology, thought about it as a game in which household members (usually spouses) are engaged with the ultimate goal of maximising their own utility. This has been a major breakthrough particularly for feminist economists because it allows for the analysis of gender inequality in a life-cycle context (Lundberg, 2001). Chiappori et al. (1988)'s collective rationality incorporates the fact that many household members care about each other by allowing utility functions to depend on the utility of other household members, with members pooling income according to a 'sharing rule'. McElroy and Horney (1981) set up a co-operative Nash bargaining model using divorce as the threat point; and Lundberg and Pollak (1993) consider the case where non-cooperation and confinement to traditional gender roles ('separate spheres') is the threat.

These threats draw their power from the deleterious effects household specialisation (more specifically, motherhood according to Lundberg (2001)) has on female bargaining power. Bargaining power is modulated via labour market outcomes, and options in the re-marriage market which we know to be gendered.[1] By specialising in household production or the 'household sphere', women's potential market earnings are undermined by lost experience and job instability owed to forced labour market withdrawal to care for children.[2] This makes it harder for women to support themselves, as well as the children for whom women usually have primary care status, in the event of divorce (Lundberg, 2001). The consequence is to undermine female bargaining power both within existing marriages and in the remarriage market. Almost all of these models are, thus, able to describe female disadvantage as originating from gender specialisation in the household (Lundberg, 2001). How pervasive gender specialisation emerges and persists in equilibrium is then an important question considered by feminist economists, endogenous with social norms and the form of State family policy (e.g. day care) (Folbre, 1995).

Enduring norms and power differentials surrounding gender makes it one important institution to consider, but there are other institutional aspects to the household which help us delve into the black-box of what economists call 'preferences for privacy'. Preferences for privacy are not simply about

[1] Chiappori et al. (1988)'s sharing rule is similarly dependent on each partner's 'outside options' in the labour or re-marriage market.

[2] Lost experience and employers discriminating against women on the basis that their job tenure will be unstable is how Lundberg (2001) explains female labour market disadvantage as originating from the household. Other explanations less directly tied to the household, such as sexism about female competence, may also exist.

wanting privacy for the self, but can also relate to a private space for someone to live with loved ones, such as a spouse or children, and engage in a 'shared life project' (Fafchamps and Quisumbing, 2007). These preferences are mediated by social norms governing who is more or less entitled to such privacy and when they may have it. Relevant institutions to consider are marriage, gender, and family. A useful expression of these institutions comes from the family sociologist Frances Goldscheider (1997), originating from the Functional sociological perspective on the family. Goldscheider (1997) describes there being two dimensions to the structure of the nuclear family. The first of these is the parent-child axis which shapes the launch of the family but also any obligations between parents and children over the life-course. The second dimension is the male-female axis which shapes the destination of the family and relates to union formation and gender roles.

In sum, there are some core economic ideas expressed in most economic models of the household. People organise themselves into households to share, produce and consume resources, and to raise children (Fafchamps and Quisumbing, 2007). Household members share and consume goods and services that are public (e.g. housing, food preparation) and private (e.g. clothing). The size of the household mediates the cost of household public goods via economies of scale: the cost of household public goods becomes cheaper when spread across more household members. However, increasing household size starts to incur a cost in terms of privacy or congestion. Economic thinking therefore provides us with three key margins that would motivate people to form their own households: increasing income (falling costs), decreasing specialisation within the household, and stronger preferences for privacy or autonomy. Since access to income, household specialisation, and social norms about autonomy are all gendered, these margins will interact in important ways with gender relations.

2.1.2 What has happened to households over time?

Census data shows household size has been declining in industrialised nations for centuries (Bradbury et al., 2014; Salcedo et al., 2012). The decline sped up in the 1890s, when industrialised countries saw household size drop from about 5 people at the end of the nineteenth century to about 2-3 people at the turn of the 21st century. Household size has also been declining in developing countries from at least the 20th century which is when data for these regions begins. Households are shrinking in the developing world from what appears to be a higher base (some countries have average household size of 8 or ten people), at a much faster pace than ever observed in developed nations, and more erratically (trends reverse in some countries), with the decline appearing to speed up in the 1980s (Bradbury et al., 2014). By the 1990s, average household size was about 5 people for developing countries (Bongaarts, 2001), and this reduced to about 3-4 people in Central and South America and East Asia by the following decade (UN, 2017).

Part of this change is the growth of single-person households which is one of the most important demographic phenomena of the modern age (Klinenberg, 2013). Living alone started trending up around the 1960s and is today at historically unprecedented levels, after being relatively rare even in industrialised nations in the first half of the 20th century (Snell, 2017). By 2018, the proportion of households that were single-person exceeded 40% in Scandinavia and Germany; reached about 35% in the Netherlands, France and Japan; and ranged between 22-28% in the United States, Spain, New Zealand and Portugal (Ortiz-Ospina, 2019). The growth and consequences of large proportions of people living alone has been the

inspiration for vigorous research activity across a range of disciplines. For example, on concerns about loneliness and social isolation in psychology and public health (Kato et al., 2017; Snell, 2017; Pantell et al., 2013); the dietary patterns of people who live alone in nutrition (Hanna and Collins, 2015); how more single-person households change demand for commuting services in urban policy (Sung, 2020); and concerns for a growing proportion of the aged who are living on their own in social policy (Portacolone et al., 2019). Solo-living has not upended familial living in developing countries to the same extent and in the least-developed regions of Africa and Asia, living alone is still almost non-existent (Esteve et al., 2020). But in middle-income countries in Latin America, Africa and Asia, there is an increase in living alone amongst men and older women (Esteve et al., 2020).

The upshot of this trend towards smaller household size is that there are many more households. United Nations (UN) (2007) estimates and projections place world household growth at 1.6-1.8 times faster than world population growth over the period 2000-2020, and happening fastest in developing countries (Bradbury et al., 2014). The UN (2007) estimated that there were 1.5 billion households in the world for 6 billion people in 2000 and projected this to change to 2.3 billion households for 7.5 billion people in 2020.[3] As households have proliferated, more women are also heading households than before, most commonly in Europe and North America (with a median share of 37 and 47% of households, respectively) than Asia and Africa (with respective medians of 19 and 27%) (UN, 2017).

2.1.3 What is driving household proliferation and shrinking household size?

The most immediate reason households are shrinking is the global decline in fertility (Bongaarts, 2001). Since 1950, world fertility levels have fallen from 4.9 to 2.6 globally, although they remain high in Sub-Saharan Africa especially (Bradbury et al., 2014). Fewer households worldwide including children is the most direct driver of this change; however, households have also seen a decline in the number of adults pointing to other forces at play (Bradbury et al., 2014). In the 1960s, sociologists Goode (1963) and later Burch and Matthews (1987) and McDonald (1992) described convergence theory. This was the theory that as countries industrialised and urbanised, household size would decrease and extended family styles would converge on the nuclear household format. This is partly the function of transitioning from an agrarian to an industrialised economy, reducing the need for extended family members to provide labour for the household's subsistence production. This also relates to the expansion of education and increasing share of higher-productivity jobs in the occupational mix (i.e. in manufacturing and then services), which made nuclear household forms more feasible. These multi-faceted processes can be well-understood by the logic of rising real income and falling costs of living. New technologies weakened the need for specialisation within households and also acted as time and cost-savers for the production of household goods (Greenwood et al., 2005). The declining share of expenditure on household public goods and concomitant rise in real income is estimated to explain 37% of the drop in the number of adults per household and 16% of the drop in children between 1850 and 2000 in the United States (Salcedo et al., 2012).

Goode (1963) also acknowledged there was an ideological angle to this change - he described the nuclear family as representing freedom and change from traditional systems. Indeed, numerous demographic and

[3]More recent estimates place the world population at 7.8 billion in 2020 (UNFPA, 2019). This means that although we were unable to locate an updated number of total households, 2.3 billion is probably an underestimate.

social changes were also taking off at his time of writing in the developed world that have come to be associated with the ideational change of the Second Demographic Transition (SDT) (Lesthaeghe, 2010). Rising education levels amongst women in the middle of the 20th century meant that in the latter half women increasingly entered the labour force. At the same time, age at first marriage began to climb along with rates of divorce, and non-marital cohabitation and childbirth. These changes have had profound implications for the form of the household in the West and elsewhere (Esping-Andersen and Billari, 2015; Lesthaeghe, 2010). Raymo (2015) found patterns of delayed marriage in Japan fully accounted for the rise in single-living between 1985 and 2010, and Bongaarts (2001) linked rising age at first marriage to shrinking household size in the developing world.

Initially, researchers debated whether the changes in marital patterns in particular were owed to increasing economic vulnerability and what is termed the 'pattern of disadvantage' or to ideational change to post-materialist values that increasingly favoured the Individual over the Family (Esping-Andersen and Billari, 2015; Lesthaeghe, 2010). The acceleration in the rate of household expansion in the developing world in the 1980s coincided with the growing momentum of globalisation and technological advances that together have made economic life increasingly precarious.[4] Precarity and uncertainty make it harder for individuals to commit fully to any living arrangement making it harder for them to achieve Pareto optimally in the long run (Baland and Ziparo, 2017). Thus, the 'pattern of disadvantage' narrative suggested that rising economic precarity undermined the ability of many people to make the financial commitment that comes with marriage - leading to a rise in cohabitation amongst the less educated. Elsewhere cohabitation rose most quickly amongst the most educated, suggesting ideational change related to the SDT was the stimulus. Since then, it appears that both of these explanations are salient to varying degrees in different contexts, again underlining the complexity of these social processes (Lesthaeghe, 2020).

What is uncontested, however, is that these changes have brought about structural change in gender relations. This is expected given how important economic theory tells us the household is for structuring relationships between men and women. Sociologists Esping-Andersen and Billari (2015) argue that trends towards 'less family' observed today are the outcome of the first half of the 'gender revolution'. This is represented by women entering the labour market manifesting a key change in gender specialisation resulting in less marriage and fertility as the opportunity cost of her labour market participation rises. Minor bounce-backs in fertility and marriage detected by Esping-Andersen and Billari (2015) in some gender-progressive Western countries they hypothesise is owed to the beginning of the second half of the gender revolution. This is when men re-enter the private sphere of the household by taking up more traditionally female work like childcare and housework to further erode gender specialisation. In this way, Esping-Andersen and Billari (2015) place structural change in gender relations at the centre of family and household change.

As such, these changes represent a reconfiguration of the male-female axis, but the parent-child axis has also been impacted. Increased individualism has been made possible in some senses by the spread of social democracies in the West that provide a social security net other than that provided by families. A result of this has been the diminishing of obligations between adult parents and their children as public and private spending on pensions has grown (Goldscheider, 1997), leading to more elderly people living

[4]Kalleberg and Vallas (2018) describes the 'four horsemen of precarious work' as originating in the 80s: digitalisation, deunionisation, financialisation, and globalisation.

alone. This has not necessarily been mirrored in the developing world where consanguineous family relations prevail and social protection is less comprehensive, keeping the tie between adult children and their elderly parents intact (Reher and Requena, 2018).

This highlights an important point: household size in the developing world is 'following suit' and decreasing as it is in the developed, but it would be unwise to uncritically apply the lessons from the latter onto the former. Structural differences in demographics, labour markets, family ties and social norms, social protection, and economic circumstances between developed and developing nations mean while the variables driving household change may be the same (e.g. union formation) the outcomes and form they take may vary. We already know that smaller households do not necessarily mean nuclear households, at least not in South Africa (Wittenberg et al., 2017), challenging mainstream ideas about convergence theory. More women heading households in Africa and Asia has coincided with more than half of them being lone-parents, unlike in North America and Europe where only about 10% of female-headed households are lone-mothers (UN, 2017). More female headship in the developing world might have increased female self-determination in some senses, but increased female disadvantage in others.

In summary, since at least the beginning of the 20th century, income has increased, the costs of living have fallen, gendered specialisation has become less concentrated, and preferences for privacy and autonomy appear to be getting stronger. Economic thinking can explain how all of these changes drive people to form more and smaller households. The trajectories of these social processes mean that households could continue to multiply long after population growth has stabilised. Although after some time lag, this could be expected to reverse if fertility also continues to decline (Bongaarts, 2001; Bradbury et al., 2014). This proliferation has wide-ranging implications for the welfare and living arrangements of families, children, women, and our environmental impact as a species. Gaps remain in our understanding of household change. In particular, changes observed in the developing world have sometimes been in contrast with what mainstream theories would suggest. This is partly owed to the diversity of the developing world, meaning that although there is value in comparative work, country case studies are still indispensable by allowing researchers to more thoroughly contextualise their findings. This leads us into a detailed background about South Africa in the next section.

2.2 Households in South Africa

2.2.1 Some historical context

South Africa is an upper middle-income country that nevertheless suffers from high levels of poverty and income inequality that is amongst the worst in the world (Alvaredo et al., 2018). This income inequality is mainly propped up by a highly exclusive labour market that still reflects historical disadvantage for Black populations and women encoded during the apartheid era. Grand apartheid - the 56 years of rule by the white minority National Party, beginning in 1948 - was the formalisation of what had been more than 300 years of racial oppression and exploitation (Smith, 2003). After decades of violent resistance, the African National Congress was voted in as the ruling party in South Africa's historic first democratic election in 1994, ushering in the democratic era and human rights and political freedom were extended to all. This landmark institutional change had important implications for household change. As Lemon

(1991, p 1) says of the apartheid era: "No other country, certainly, has embarked on so thorough a reorganisation of its urban space for the purposes of segregation". The same could easily be said of the country's rural space.

The apartheid economy was based on the concept of a rural cheap Black African work force supplying labour to the white economic centres in the cities and the mines in proportion to their needs (Lemon, 1991). As part of this scheme, Black populations were restricted to low-skilled, low-paying jobs (the 'colour bar') while skilled jobs were reserved for whites (Nattrass and Seekings, 2010). The schooling system was also segregated with the much-underfunded Black schooling system aiming to produce a class of compliant labourers more than to educate, and provided only basic literacy and numeracy training (van der Berg, 2007). These policies, and racial segregation more broadly, were ideologically based on the notion of 'separate development' of the race groups (Wilson, 2001). The Bantu Authorities Act of 1951 set up ten homelands - four of which were considered fully 'independent' and excluded from censuses - in which Black African South Africans were to live. The homeland system further reinforced the flow of circular labour migrants between the mines and their family in the rural homeland. This pattern is typified by men living for long periods of time close to their site of work, often in single-sex worker hostels, and remitting to their families in the homelands (Wilson, 2001). The Group Areas Act, also passed in 1951, segregated the cities and empowered the government to forcibly remove the urban Black (mainly Coloured and Indian/Asian) populations living in the center of cities to the peripheries. The segregation was highly successful; by 1985, only about 10 percent of the urban population lived outside of its designated group area (Lemon, 1991). No less than 99.7% of whites already lived in what became white group areas. As a result, the apartheid state heavily intervened in the living arrangements of all Black South Africans, whilst leaving the living conditions of white South Africans intact and undisturbed.

The clear rural-urban divide envisioned by the planners of the apartheid economy was unsustainable both because the economy needed a settled urban labour supply and because Black Africans themselves wanted to be closer to cities where most economic opportunities were (Smith, 2003). The latter was the case because of the close-to-zero employment opportunities in many of the homelands combined with the increased land dispossession of Black farmers outside of the homelands (Lemon, 1991). Black urbanisation was a feature of South Africa in the 1960s and especially the 1970s (Smith, 2003) until the system of segregation collapsed in 1986 with the abolition of the pass system. The first democratic election was held eight years later. In the decade leading up to the election, the mining industry which was a key employer of Black African men, underwent a series of crises including a plummet in the price of gold instigating a contraction from which the industry would never recover. For these and other reasons, the mines formalised their workforce to a smaller corps of 'career miners' leaving many of the men who had effectively shared these jobs between periods of time at home, without work (Nattrass and Seekings, 2010; Wilson, 2001). The result was that unemployment amongst the Black African population started climbing quickly in the 1970s but especially so in the early 1990s when the mines started shedding jobs and has essentially never subsided. When the election happened in 1994, unemployment was a already a serious concern.

The apartheid system systematically impoverished and undereducated Black South Africans and interfered in family set-up through the "long dislocation between residence and workplace" (Smith, 2003, p 8). The result was that in 1994, at the end of apartheid, unemployment was high and marriage rates were low amongst Black African South Africans in particular. The migrant labour system and

land dispossession of the apartheid era together engineered South Africans to be reliant on wage-work as a livelihood strategy. This has had disastrous effects on the lives of many South Africans as the post-apartheid economy struggled to provide jobs, resulting in high numbers of discouraged workseekers. Race and gender continue to be markers of disadvantage in the labour market: male and female strict unemployment rates increased from 17 and 25%, respectively, in 1994 to 26.8 and 31.5% in 2019 (Casale et al., forthcoming). Income was highly concentrated in the hands of whites at the end of apartheid making South Africa one of the most unequal countries in the world with a household income Gini coefficient of 0.68 in 1993; not much changed 21 years later when the Gini remained high at 0.66 (StatsSA, 2019).

After 1994, a liberal constitution was drawn up and the homelands were dissolved. From the late 1990s, welfare spending expanded considerably as the new government sought to uplift basic living standards. Social grants targeted mainly at children and the aged can be credited for a large portion of the reduction of income poverty over the period (Leibbrandt et al., 2010), despite the notable omission of an unemployment grant. Access to free schooling, housing, and a base level of electricity and water expanded substantially over the period meaning multi-dimensional poverty dropped more drastically than money-metric poverty (Finn et al., 2013). Although the free provision of housing and services can be expected to considerably reduce the cost of moving out of home, implementation has been beset by procedural backlogs and quality control issues (National Treasury, 2004).

2.2.2 Households in post-apartheid South Africa

After apartheid ended and restrictive laws governing where people could live and work were lifted, many researchers wondered how marriage, households and migrancy would evolve in post-apartheid South Africa. For example, would temporary circular labour migration give way to more permanent migration where working men and their families could settle together in urban areas (Posel, 2010)? And, would this allow marriage to 'bounce back' by allowing men and women to live in the same locations (Hosegood et al., 2009)? And, ultimately, what would this mean for households and household composition (Russell, 2003; Amoateng, 1997; Posel and Hall, forthcoming)? The post-apartheid period also brought with it the onset of regularly collected household survey data with a sampling frame including the populations living in the former homelands from StatsSA as one tool researchers could use to answer these questions.[5]

Efforts to use these data to answer some of these questions has been somewhat frustrated by how the household is expressed in the data. As mentioned in Chapter 1, the surveys capture immediate living arrangements by defining a household member as someone who has on average slept four nights a week in that household in the four weeks prior to enumeration. This definition is challenged by patterns of oscillating labour migration whereby the membership of a rural household changes in correspondence with whether a migrant member is at home or away for extended periods to work and remit earnings on which the rural household depends (Spiegel et al., 1996). To the extent that households are spatially confined and defined as 'eating from a common pot', the urban migrant and rural homestead could be considered two households when he is away. But inasmuch as the members are all acting in common economic purpose to share income, the household is stretched across space in what is now a well-established notion in the South African literature of a 'stretched household' (Spiegel et al., 1996).

The narrow household definition in StatsSA survey data precludes the ability to identify stretched

[5]Prior to this, census data excluded large portions of the Black population by excluding the four independent homelands.

households by not recording absent household members in the household roster (Posel, 2003). Outside of the roster, some StatsSA surveys[6] include a separate module for migrants specifically, although the questions in these modules have not always been consistent over time (Posel, 2010). Researchers have usually relied on other sources to show that internal labour migration continues to be relevant in post-apartheid South Africa, with rural livelihoods still shaped in a major way by their migrating members (Collinson, 2010; Posel, 2010). However, the conventional form of migration - that of a man engaged in circular movement between often a worker hostel on a mine and a rural homestead - is changing. More women started migrating with the wave of increased female labour market participation that happened in the 1990s (Posel and Casale, 2003), but by 2008 it appeared that migration was decreasing or at least giving way to more permanent settlement (Posel, 2010). Recent evidence shows that migration then ticked up again after 2010, as the country started to emerge from the recession induced by the 2008 Global Financial Crisis (Posel, 2020).

Migration is intertwined with the inexorable urbanisation of the South African population, in general, which has been happening since at least the 1970s, and likely earlier (Parnell and Crankshaw, 2013; Mabin, 2003). The average South African today is an urban-dweller, and notably so is the average Black African South African, despite the segregationist designs of the apartheid era. Rural Black Africans outnumbered their urban counterparts by two to one in 1995, but the share of Black South Africans living in an urban area crept over the halfway mark for the first time in 2005 and has continued to inch up since then (own calculations using the GHS). Much of the rural Black African population are children left under care of other kin while parents migrate to work in urban areas where the lion's share of economic opportunities exist (Hunter and Posel, 2012). This complicates the distinction between urbanisation and labour migration. Although movement to the cities is becoming more permanent, many Black Africans remain 'double-rooted' (Bank et al., 2020): living mostly in urban areas to earn their livelihood but maintaining family and cultural ties in rural areas. This is represented by ongoing investment in rural areas (Bank, 2020) and many workers returning to rural areas upon retirement or illness (Clark et al., 2007).

Marital rates have been declining since at least the 1950s for all groups of South Africans (Garenne, 2016), but most swiftly for Black Africans and Coloureds as apartheid policies which separated men and women in the labour migration system contributed to the breakdown of family formation (Hunter, 2010; Smit, 2001). The decline in general and divergence by race has continued into the post-apartheid period so that by 2010, Black African women were 40 percentage points less likely than white women to have ever married (Posel and Rudwick, 2013). This divergence has been linked to much higher rates of unemployment amongst Black African compared to white men (Posel and Casale, 2013), undermining their marriageability and making the financial requirements of bridewealth increasingly unattainable (Posel and Rudwick, 2014). As marriage has declined, childbearing has largely been de-coupled from marriage (Hosegood et al., 2009): an astonishing 68.2% of children under 15 years lived without co-resident father in 2014 (Hall and Posel, 2019). Cohabitation has also not increased to compensate for the decline in marital rates and divorce rates remain low (Posel and Rudwick, 2013) meaning that in South Africa men and women are increasingly living separately, in a pattern of household gender polarisation we revisit shortly (Posel and Hall, forthcoming).

The trajectories of marriage and migration have had important implications for households and their

[6]These are the 1996-9 OHSs and the 2002-5 Labour Force Surveys (LFSs).

composition. Historically, Black African households in South Africa have followed a consanguineal and patrilineal system (Siqwana-Ndulo, 1998; Russell, 2003) rather than the conjugal system associated with the West and the nuclear family. The primary unit of the conjugal system is the married couple who may go on to raise children and households only include these immediate family members. In the consanguineal system, the primary organising principal is the kinship network and common ancestry. Wives often move into the household of her husband's family and households include extended family members, resulting in household structure and size that is complex and fluid (Russell, 2003; Seekings, 2008).

Having noted the decline in average household size and fertility at the beginning of the post-apartheid period, some scholars wondered whether Black African households were becoming more nuclear, resulting in one of the central debates in the South African household literature (cf. Amoateng (1997); Steyn (1995); Russell (1994, 1998); Ziehl (2001, 2002)). Consensus was difficult to reach in part because researchers often used the term 'nuclear' to mean different things (Ziehl, 2002; Russell, 1998). The term nuclear *household structure* is often used to mean only immediate family members are present in the household (spouses and children) at a particular point in time, like in a snapshot provided by a household survey. This is different to a nuclear *family pattern* which refers to the rules people apply to household formation over the lifetime based on their ideals of family formation (Seekings, 2008). Nuclear household *structures* can crop up even in the consanguineal extended *family pattern*, leading Ziehl (2002) to caution against counting up nuclear households in a survey as evidence of nuclear family patterns.

Consensus in this debate was also difficult to reach because the evidence this literature often drew on was based on small localised studies with different definitions of a household (Ziehl, 2001; Seekings, 2008). As time has passed and more data (with an invariant definition of a household) has become available, it appears more likely that declining household size is an aggregate outcome belying increasingly complex compositional shifts within households. For one, households have undergone a process of household gender polarisation whereby men and women increasingly live in male- or female-only households: in 1995, 26% of households were home to adults of one gender only and by 2018, this had risen to 46% (Posel and Hall, forthcoming). For women, this usually still involves children, but for men this has been realised in an astronomic rise in single-person living. Dovetailing with this process is one of household employment polarisation whereby some (usually smaller) households are home to fully employed adults whereas other (usually larger) households are supporting many unemployed adults (Pirouz, 2005; Wittenberg, 2017). In the absence of State-sponsored unemployment support, most of South Africa's large unemployed population have been shown to 'stay at home' and draw on the household as a social security net (Klasen and Woolard, 2009; Ebrahim et al., 2013; Keller, 2004).

As such, South African household composition could be described as growing at the 'poles' of complexity. On the one hand, the share of single-person households has mushroomed from 17 to 26% of the household stock between the 1996 and 2011 censuses (own calculations). On the other hand, researchers using more recent data have repeatedly confirmed the continued relevance of complex households structures (Wittenberg et al., 2017; Posel and Hall, forthcoming). An outcome of these compositional shifts is that female-headed households are increasingly reliant on grant income and female labour market earnings (Rogan, 2013; Posel and Rogan, 2012), which remain lower than male earnings (Mosomi, 2019). This change is identified as a main reason behind the widening poverty risk differential between male- and female-headed households, where female-headed households continue to face the higher risk of falling into poverty (Rogan, 2013; Posel and Rogan, 2012).

The State's provision of free housing and services aimed to reduce poverty, but has been shown to interact with household formation in complex ways. Wittenberg and Collinson (2020) show that rapid household proliferation in the rural Agincourt area of Mpumalanga was related at least in part to the provision of free housing, colloquially called 'RDP' housing. These authors also identify this outcome to be the result of new household formation, as opposed to the shrinking of existing households, for example. In related work, this process was found to be happening so quickly it was undermining the ability of the State to improve the backlog of households waiting to be connected to the electricity grid in the same area (Harris et al., 2017). Social grants have also been shown to impact household composition. In particular, the onset of pension income is associated with key compositional changes: young children move in and women of working age move out, often in search of work (Ardington et al., 2009; Ranchhod, 2017; Edmonds et al., 2005). This effect is often specifically associated with female pensioners, bringing the dimension of gender to the fore, which we turn to next.

2.2.3 Gender in post-apartheid South Africa

Household formation in the post-apartheid period cannot be properly analysed without an understanding of how patriarchal gender norms have modulated the effects of other drivers of household change, e.g. the labour market, social expenditure. Apartheid South Africa was a highly patriarchal society, intersecting with race in complex ways (Shefer, 2010). White women were more likely to take on the traditional home-maker role, whilst many Black women worked as domestic workers in their homes (Rommelspacher, 2020; Nattrass and Seekings, 2010). Black African women were treated as minors in terms of apartheid law, making them wards of their male relatives and thus barring them from owning property themselves (Nolde, 1991). After apartheid, along with expanding human rights to all South Africans, women's rights in particular were expanded in a series of legislative changes in the 1990s. The extension of equal treatment in terms of property ownership as per the new democratic constitution allowed Black African women to own property for the first time, and laid the foundation to incorporate women more formally into the labour market (Bentley, 2004). Other important augmentations of women's rights at the time include the outlawing of marital rape (Karimakwenda, 2020); legislation preventing domestic violence (Vetten, 2013); and, the legalisation of abortion (Althaus, 2000).

Despite these advances on paper, many South Africans still subscribe to patriarchal ideas about gender many years after the first democratic election (Timol et al., 2019; Boonzaier, 2005; Albertyn, 2009). In some senses this is aggravated by women remaining in a subordinate position to men in customary law in South Africa, creating tension regarding women's status in law with implications for women's *de facto* position in society (Venter, 1995; Bentley, 2004). These norms feed into South Africa being one of the most unsafe countries for a woman to live. South Africa has the highest rate in the world of women being killed by a current or former intimate partner (Abrahams et al., 2009); and 70% of female respondents in one recent national survey reported experiencing some form of emotional abuse by their intimate partner in their lifetime (Zembe and Adjiwanou, 2019).

An analysis of attitudinal data from 2012 found that, although South Africans supported women working, their role as a care-giver to children came first. Men's role in the family was associated with provision and earning: 45% of women and 48% of men agreed that 'a man's job is to earn money, and a woman's job is to look after the home and family'. Evidence that attitudes are becoming more

progressive over time is limited (Timol et al., 2019). Fatherhood in South Africa has long been narrowly defined as provision in findings that cut across socio-economic status, education level, religion, and race (Makusha et al., 2019; Rabe, 2006). This interacts with the country's high level of unemployment, with unemployed men repeatedly reporting feelings of shame and alienation that they cannot provide for their families (Makusha et al., 2019). The association between fatherhood and provision has been reinforced by labour migration whereby "fatherhood is enacted within the constraints of paid employment" (Makusha et al., 2019, 210).

Although men may be away from their families precisely so that they can provide (Madhavan et al., 2008), nationally-representative time-use data and surveys asking about primary care responsibility show that the overwhelming majority of childcare is done by women even when men are co-resident (Hatch and Posel, 2018). Further, this care imbalance is not compensated for by a similar imbalance in financial provision. Although men are sometimes more likely to take on certain costs (e.g. schooling), women are found to be responsible for children both in terms of primary care and financial support in the majority of cases (Hatch and Posel, 2018). The persistence of these gendered divisions of labour come to bear on the trajectory of change in household formation.

Overall, there has been a lot of work conducted on the household, its form and its wider implications in South Africa. The work this thesis contributes is firstly about measurement. As mentioned, some of the first debates around household change in South Africa were hampered by researchers drawing on different data sources with different definitions of a household (Ziehl, 2001; Seekings, 2008). Even though we now have more than two decades of data collected by StatsSA with an invariant definition of a household, data quality issues still undermine the accurate study of household change and this is one issue we seek to address in Part I. Part II then addresses the next logical question, which is how do trends in known drivers of household change cohere with trends in household formation? Better measurement of single-person households in particular is relevant here in allowing us to more reliably trace trends in living alone than has previously been done, including for considering the implications of single-person households for gender polarisation and household composition. Part III investigates trends in leaving the parental home in South Africa and how this overlaps with household formation, a subject that has received little scholarly attention using survey data. In Part IV, we reflect on the meaning of our findings for some of the themes brought up in this review; such as, household composition, gender polarisation, living alone, and the debate about nuclear households. The patterns of household formation and composition that emerge in South Africa connect to the broader social forces that swept through the developed world, but are also shaped in critical ways by the country's unique history. The extent to which global social forces explain or are subsumed by South Africa's context provide important lessons for understanding household formation more generally.

Part I

Measurement

Chapter 3

Reweighting South African household survey data to restore representativeness

3.1 Introduction

The empirical work in this thesis is based on a series of national household surveys from StatsSA, which we stack together and refer to as the OHS-GHS series. In this chapter, we introduce these data and cover the key data quality issue we need to address in order to use the series to extract reliable trends on household formation: the survey weights. Survey weights calibrated after sampling are one of the main tools survey statisticians use to credibly call national household survey data 'nationally representative'. Without this step, the survey data could look very different to the population it was designed to represent, due to a range of issues in the design and sampling stages. A reliable set of survey weights is then critical to the extent that researchers want to extract estimates and totals reflective of South Africa as a whole, a quality important to many researchers and certainly to the ability to track the progress of the post-apartheid project, in general. This quality is also central to the goal of this thesis to track change in post-apartheid household formation, for which the extraction of accurate trends in household and person counts is of prime importance.

Although calibration of survey weights is used to fix deficiencies in representativeness, if done poorly, it can create more problems than it solves (Smith, 1991). This is the case in the StatsSA data we use to create the OHS-GHS series, in which an unusual survey weighting practise artificially detaches people from the households in which they live, leading to a series of statistical and conceptual problems compromising the quality of the data. This happens because weighting practise in the publicly released StatsSA data does not follow sampling practise, when this should always be the case. Sampling practise in South African national survey data is that every person in a sampled household is surveyed. The

implication is that weights should be 'integrated', meaning members of the same household have the same weight, and that there should be no distinction between person and household weights because the chance a person will be sampled equals the chance their household is sampled. In other words, there should be one weight in the data, equal within households.

A subset of household surveys, including the October Household Survey (OHS) and the General Household Survey (GHS) we use for the OHS-GHS series, collect information at both the person and household level, making both of these elements important units of analysis. Curating a consistent and reliable series of population and household counts over time is, therefore, a first-order output of the survey data. The data is usually released in two files, a person and household file each including a weight. If sampling practise was followed, the weight for a given household in the household file should equal that for a member of the same household in the person file. However, the original series of integrated weights in the GHS calibrated only using person information did not produce a realistic series of household counts required, in particular, by policymakers. The need for an accurate series of household counts combined with the technical difficulty of adding household information to the calibration model prompted the separation of the calibration for these main units of analysis into two parallel and mutually exclusive processes in 2009. The result was that the household weight in the household file and the person weight in the person file are are calibrated on completely different information and never equal to each other.

This practise contravenes sampling practise and yields two major problems: inconsistency is introduced into estimation since multiple weights lead to multiple estimates of the same statistic (including total person and household counts); and, people and households are incoherently de-linked from each other conceptually. This happens because researchers have to choose between weights that either yield a representative person universe or a representative household universe, but not both at the same time. The result is a needless reduction in how representative the data is of the world and in the number of research questions this data can be used to answer in a conceptually coherent way. Since this procedure is embedded in more than 20 years worth of data, such an erosion of data quality counts as a considerable cost both to the country's research agenda and to the efficiency of public spending. It also means it is not viable for us to use the weights StatsSA released with the original source data in our OHS-GHS series.

Fortunately, these costs are not incontrovertible. The nature of the calibration process, as well as, the fact that it is undertaken after a survey is collected means it can be remedied relatively painlessly compared to other influences on data quality, like survey design or sampling errors. In this chapter, we investigate whether we can combine the information StatsSA uses in two separate calibration processes, into one procedure and reweight the data. As we explain over the course of the chapter, it is not apparent at the outset that this could practically be done because, as far as we know, StatsSA have never released survey weights calibrated using both auxiliary person and household information at the same time. We use cross-entropy estimation to successfully recalibrate the survey weights for the OHS-GHS series in a way that coheres with the survey sampling practise. The new cross-entropy weights are able to restore the mutual representativeness of people and households, and thereby recover the number of questions these data can be coherently employed to answer. An immediate output is a new series of integrated cross-entropy weights for the OHS-GHS series that yields person and household counts that conform with census benchmarks.

In the next section we briefly review the motivation and mechanics of survey weight calibration. Following this, we describe the challenges facing researchers who want to count the total number of

households and people in South Africa using the available survey data in Section 3.3. In this section, we also introduce the source data for the OHS-GHS series in more detail; document the history of weight calibration in these data; and, explain the problems caused by the StatsSA practise. Section 3.4 details the cross-entropy technique and how we combine person and household information into a single calibration model. Section 3.5 presents our results, both in terms of the validity of the calibration itself, as well as, how our new weight solves the problems created by the StatsSA practise. In Section 3.5.3, we test the performance of the new weight on statistics other than household and person counts, and present caveats on the scope of our improvements to the weighting system. Section 3.6 concludes.

3.2 Setting up quality household survey data

The best type of survey data for collecting information about people and households is a census since this surveys the entire population. However, censuses are very expensive operations and entail large-scale complicated logistics so that they are not usually feasible on a regular-enough basis for the type of monitoring household surveys are often used for. This is especially the case in developing countries with capacity constraints on their statistics bureaus. It is much more feasible to survey a sample of the population. This smaller operation is more cost effective, can be carried out more regularly, and allows for more detailed and directed questionnaires.

The danger with this approach as opposed to a census is that the sample collected can end up looking quite different to the actual population because of non-response, luck surrounding the drawing of a single random sample, and other measurement error occurring during the surveying process. Crucially, with a combination of prudent survey design and careful survey weighting, statistics estimated from the sample will satisfactorily approximate what they would have been had they been estimated from a census (Deaton, 1997). Ensuring that the sample is representative of the population, or sampling frame, has received a lot of scholarly attention in recent decades and is the topic of an extensive academic and practical literature (Deaton, 1997; Lavallée and Beaumont, 2015; Deville and Särndal, 1992).

Survey weights are an essential aspect of representativeness. Survey weights are scaling factors assigned to each unit of observation to make the sample representative of the population. There are usually three steps to this process: (1) assigning the design weights which account for probability of a unit being selected for sampling, dependent on survey design; (2) adjusting the weights to compensate for units that were selected to be surveyed but were not surveyed for some reason (e.g. non-response; non-coverage); and (3) tuning the weights so that estimates from the weighted sample reflect known population totals, often taken from a census, in a process known as calibration (Lavallée and Beaumont, 2015).

Common survey designs, such as two-stage sampling and stratification, usually mean the probability that a given unit will be selected for the sample varies from unit to unit (Deaton, 1997). This means that each unit in the sample represents a different number of units in the population and thus the sample needs to be weighted accordingly to achieve unbiased estimates of population means. This initial weighting is the role of the design weights, which are defined as the inverse of the probability of a unit being selected for sampling, and, when summed, will yield the population total. Most national surveys collect information at the household and person level, which means that the probability that a person is selected for sampling is conditional on the probability that their household is selected for sampling.

Weights additionally need to be adjusted for accidental ex-post survey factors (Lavallée and Beaumont, 2015). Invariably, the actual sample collected differs from that which was designed due to participant non-response, non-coverage, or other measurement error. As such, the probability of inclusion will vary across sample units for reasons other than design. The survey statistician models non-response in order to make this adjustment. Non-response adjustment is less rigorous than the design weight since modelling decisions by the survey statistician are embedded in the weight.

The third stage of weighting, with which this chapter is most concerned, is calibration. Calibration is the process of weighting the sample so that the estimates conform with known population totals from an auxiliary data base, often demographic characteristics about age, gender and geography from a census (Lavallée and Beaumont, 2015).[1] The statistical problem of calibration is that unnecessary randomness is introduced into estimation by the process of sampling and sampling design. The aim is to improve the quality of inferences in case a poor sample is selected by weighting the sample to make it as representative as possible of the true population.

Key work in this field comes from Deville (2000) and Deville and Särndal (1992) who explain how calibrated weights, w_i, can be solved for by minimising a distance function between estimated population totals using the design weights and known population totals from an auxiliary data source. Let x_i be the value of a variable for the ith observation in a sample s of size n drawn from population size N. The population total for $x = (x_1, ..., x_i, ..., x_N)$ is defined as $t_x = \sum_{i=1}^{N} x_i$. Consider now, that we know the value of t_x from some auxiliary information source like a census. An unbiased estimator of t_x would involve weighting up the observation values with the design weights, d_i, and then summing so that $\sum_{i \in s} d_i x_i = \hat{t}_{x,d}$.

If a bad sample is selected, this estimator could be very far from the actual t_x in which case calibration is motivated. The aim is to choose new calibrated weights, w_i, to be as close to the design weights as possible in order to maintain unbiasedness (Deville and Särndal, 1992), but also so that they agree with the known population totals so that:

$$\hat{t}_{x,w} = \sum_{i \in s} w_i x_i = t_x \tag{3.1}$$

Since the new weights should be kept as similar as possible to the design weights, the optimisation problem is therefore to minimise a distance function $G_i(w_i, d_i)$. That is, to minimise $\sum_{i \in s} G_i(w_i, d_i)$ subject to equation 3.1 (Särndal, 2007). This yields the following solution for the calibrated weights:

$$w_i = d_i F(q_i, x_i, \lambda) \tag{3.2}$$

where $1/q_i > 0$ is a set of known scaling factors[2] unrelated to d_i; and, λ is solved using the constraint in equation 3.1 and has a probability tending to one as $n \to \infty$.

The calibration is ultimately informed by a whole vector of known population totals of a size less than the total number of observations. A basic example is that a census could inform us about the number of women in the population, the number of people living in rural areas, as well as the number of people

[1] The design weights will not always conform with population totals because in some cases, the survey is designed to over- or under-sample certain sub-populations for cost or statistical efficiency reasons. For example, in South Africa, the Indian sub-population is very small. As such, they are usually oversampled in national household surveys in order to collect a large enough sample for robust estimation. This means that they are designed to occur in the sample at a higher rate than what they occur in the population.

[2] The scaling factors arise from the choice of $G(w, d)$. The function is chosen so that the first partial derivative with respect to w is $g(w, d) = \partial G(w, d)/\partial w = g(w/d)/q$.

aged 25-35. The calibration therefore needs to simultaneously optimise the weights for gender, region, and age, putting several restrictions on rural women aged 25-35.

As such, obtaining the calibrated weights is not as objective a process as obtaining the design weights which depend entirely on the survey design (Deaton, 1997). By contrast, calibration is essentially a modelling problem and the modelling can be of better or worse quality, depending on the standard of the auxiliary information available and the modelling decisions made by the survey statistician. If the external source providing the known population totals is of poor quality or the sampling errors from non-coverage or non-response are large and difficult to overcome, the calibration will be undermined. In these cases, calibration can create as many problems as it tries to solve (Smith, 1991).

Modelling choices include specification of the distance function, $G(w, d)$; choice of constraints from the auxiliary data; as well as, the level of disaggregation of the constraints. On the one hand, the more constraints that are used (e.g. age, region, race) and the more detailed (e.g. finer age bands, more disaggregated regions), the closer the sample should get to resembling the true population and the more precise your estimates. However, too many constraints relative to raw sample size can undermine the computing of the calibration and introduce other small sample biases as the sample is cut into finer and finer cells (Deville and Särndal, 1992).

The survey statistician thus faces a trade-off between the improvement in the accuracy of a core set of estimates and the potential distortion caused by the calibration. For example, ensuring accurate person and household counts may be prioritised, at the cost of misshaping another, less fundamental statistic. Modelling decisions by the survey statistician are thus embedded in calibrated weights and researchers may be more or less in agreement with these decisions and may want to make their own choices (Deaton, 1997). For these reasons, it is important that calibration method is well-documented and that design weights are released along with calibrated weights.

In the case of South African national household survey data, a sub-optimal calibration approach results in a very restrictive and unrealistic conceptual framework being embedded into almost 20 years worth of data and introduces unnecessary noise into estimation. National household survey data sets are not released with original design weights, meaning researchers are unable to make their own decisions about how to proceed with weighting. Researchers instead have to adopt the weights provided, unless they go through the process of calibrating their own weights, which is precisely the exercise undertaken by this chapter. The next section considers the situation facing a researcher wanting to count households and people in South Africa; introduces the OHS and GHS data along with the history of the StatsSA weighting procedure and the conceptual problems it creates.

3.3 Counting people and households in South African household survey data

The two most fundamental demographic statistics found in any national household survey data are, perhaps, the number of people and the number of households. Because they are so foundational, if the data cannot accurately represent these two counts, confidence in any analysis on more complex statistics is undermined. These two statistics are often the frame to which survey weight calibration is tethered and - if the survey data is collected on a regular basis - attention is dedicated to achieving a consistent

series of these counts. As previously discussed, the best tool for counting people and households in a country is a census. South Africa has census data for the post-apartheid period for 1996, 2001, and, 2011; as well as, a Community Survey in 2007 and 2016. Community Surveys are larger-than-usual national household surveys conducted in lieu of a census due to capacity constraints (StatsSA, 2007*b*).[3] Outside of census years, we rely on annually collected household survey data.

In Figure 3.1, we chart household and population counts using various household survey data for South Africa over the post-apartheid period from 1995 until 2015.[4] The OHS and GHS data used as the source for the OHS-GHS series is described in detail in the next sub-section. Both the OHS and the GHS are national household surveys collected by StatsSA. For comparison, we also include estimates from data either collected or curated outside of StatsSA. The Post-apartheid Labour Market Series (PALMS) is a curated series of StatsSA labour market data by Kerr et al. (2019). PALMS comprises the OHS data (1993-9); the Labour Force Surveys (LFSs) (2000-7); and the Quarterly Labour Force Surveys (QLFSs) (2008-2019). The PALMS version of these data include additional cleaning, harmonising, and notably for our purposes, separate calibration of survey weights, compared to the source data. The National Income Dynamics Study (NIDS) is the country's only nationally-representative panel survey, implemented by SALDRU (2016) since 2008.

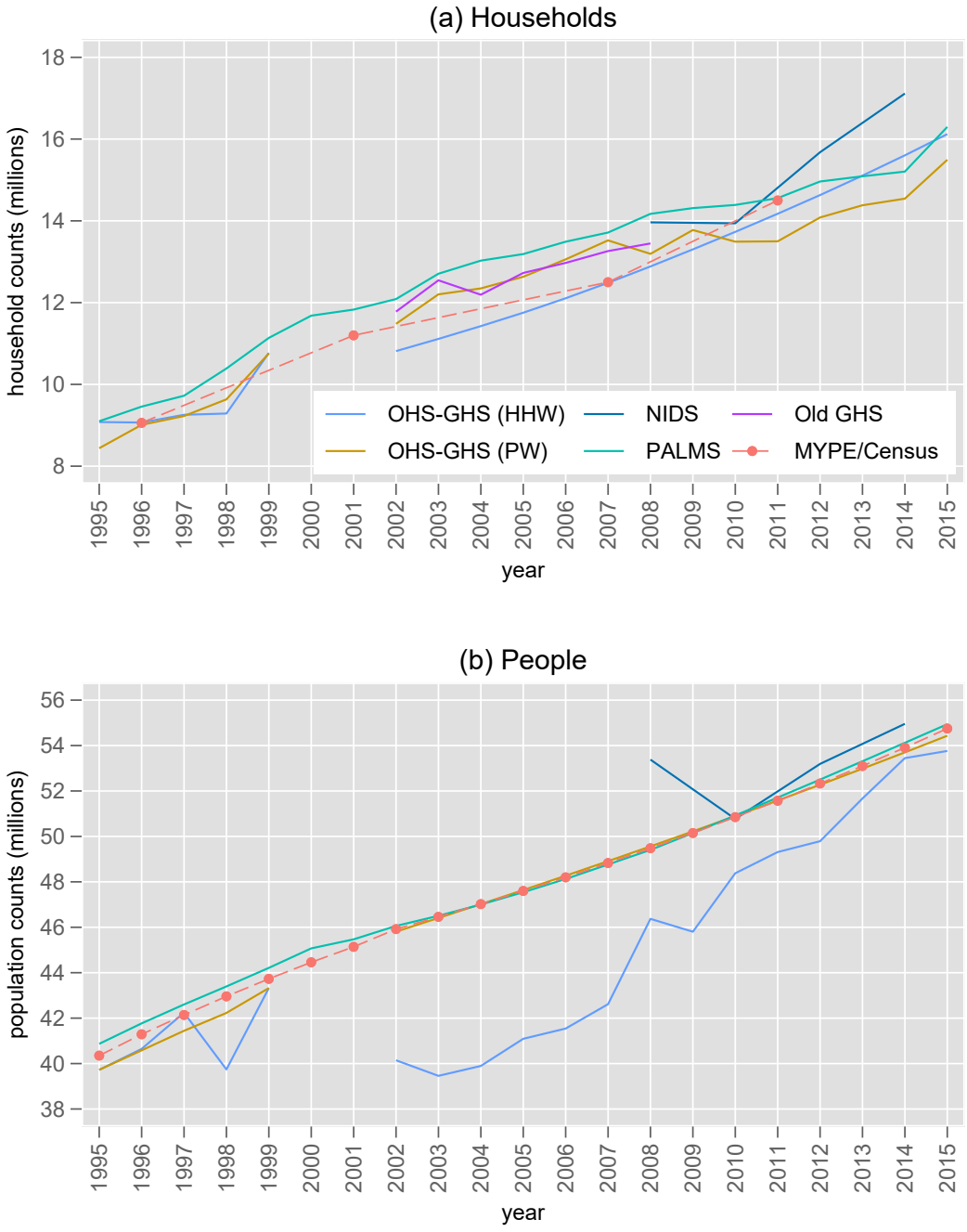
In Panel (a) on households, estimates from these surveys are compared to officially released numbers from StatsSA (2001*a*, 2012*c*), based on the census ten percent samples. The census ten percent samples are publicly available (StatsSA, 2015*a*, 2011, 2015*b*) and released with at least two files of person and household information, also each with their own weights which account for any undercount detected by a post-enumeration survey. The person file is a ten percent sample of all people present in South Africa on census night, except foreign diplomats and their families (StatsSA, 2012*a*). The household file is slightly more limited in being a ten percent sample of all households excluding institutions (e.g. prisons, hospitals) and in 1996, additionally excluded worker hostels (StatsSA, 1996), a point to which we return in Chapter 4. We also use the 2007 Community Survey data (StatsSA, 2010*a*), which has the same person-household file set-up as the censuses, with different person and household weights. In Panel (b), we compare survey estimates for the total population count to the Mid-Year Population Estimates (MYPE), which are demographic estimates released every year from StatsSA's Demography division and consolidated with a consistent underlying population model (StatsSA, 2018*a*).

Figure 3.1 shows that the census counted just over 41 million people living in 9 million households in 1996; in 2011, there were just over 51 million people living in 14.5 million households. Some of the surveys in Figure 3.1 do not conform with the census in census years; and, outside of census years there is little agreement amongst surveys about total people and household counts in a given year. The series of person counts plotted by the GHS household weights in Panel (b), performs particularly poorly, highlighting the discrepancy between the person and household level in the StatsSA weights. Overall, this is a disconcerting level of uncertainty around these two fundamental counts.

[3]The 2007 Community Survey had a sample size of 949 thousand people and 246 thousand households, making it about three times the size other StatsSA household surveys (StatsSA, 2007*a*). For comparison, in 2007 a census would have surveyed 48.8 million people by StatsSA's mid-year population estimate for that year.

[4]This time period ends in 2015 because 2015 was the most recent year of data available when the data was first downloaded from the DataFirst website in 2017. Since the end of our study period is limited to 2011 anyway for reasons explained in Section 3.4, it was not necessary to update this plot further. We also begin in 1995 and omit 1994 for similar reasons.

Figure 3.1: Total population and household counts in South Africa in various household surveys, 1995-2015



Notes: own calculations using HHW = Household weighted OHS-GHS series; PW = Person weighted OHS-GHS series; NIDS = National Income Dynamics Study; Old GHS = GHS weighted using the original integrated weight; PALMS = Post-Apartheid Labour Market Series v3.3; Census = household totals from census 1996, 2001, and 2011 and Community Survey in 2007 in Table 3.10 in StatsSA (2012*c*) and Table B in StatsSA (2001*a*); MYPE = mid-year population estimates from StatsSA (2018*a*).

3.3.1 Constructing the OHS-GHS series

Out of the set of household survey data sets collected by StatsSA, the most appropriate to use for counting people and households is the October Household Survey (OHS) and the General Household Survey (GHS). The OHS is the only large nationally representative annual household survey undertaken by StatsSA in the period 1994-1999.[5] Additionally, the OHS collects data about relationships within the household, defined in reference to the household head (e.g. spouse of the head, child of the head). After 1999, the OHS was conceptually split into the Labour Force Surveys and General Household Surveys, with the former focusing on economic outcomes and the latter on socio-demographic outcomes. The GHS - which only launched in 2002 - is therefore the survey that inherited questions about household relationships from the OHS, making it the logical choice for the latter half of the period.

Both the OHS and the GHS are cross-sectional and survey approximately 30 000 dwelling units based on about 3 000 Primary Sampling Units drawn from the Master Sample of enumerator areas used during the most recent census at the time. Exceptions are that the 1996 and 1998 October Household Surveys only surveyed about 16 000 and 20 000 dwelling units, respectively, due to budget constraints. A stratified, two-stage cluster sampling design is employed in each case, stratified at the provincial level.[6] Data is self-reported to the enumerator (or by proxy in the case of an absent respondent) and covers the spectrum from demographic and household information to basic labour market data. Between these two surveys then, we have large samples of nationally representative cross-sectional data on individuals, households, and their structures for every year in the period 1994-present, with the exception of 2000 and 2001.

We combine 20 releases of the OHS (1994-9) and GHS (2002-15) surveys (StatsSA, 2010-2013, 2011-2018*b*) into the OHS-GHS series by merging person and household information; stacking the cross-sections by year; cleaning and harmonising a subset of variables useful for our purposes; and, in particular, keeping both the household weight from the household files and the person weight from the person files. The result is a data set of about 1.8 million people and 483 thousand households spanning the period 1994-2015, with the exception of 2000 and 2001.[7] Also relevant for comparability between years is when the Master Sample changes. OHS 1994 and 1995 were based on enumeration areas from the 1991 census; OHSs 1996-1999 and GHS 2002-2004 were based on the 1996 Master Sample using the 1996 census; GHS 2005-7 were based on the 2004 Master Sample from the 2001 census; and GHS 2008-2014 were based on a Master Sample originally designed for the Quarterly Labour Force Surveys (QLFS) which replaced the LFS starting in 2008 and also based on the 2001 census (StatsSA, 2008*a*). In 2015, StatsSA started using the 2013 Master Sample based on the 2011 census. The 2008 GHS is also notable for being beset by a higher-than-usual amount of sampling errors (e.g. listing and capturing errors) so that a greater proportion of sampled dwelling units than usual were excluded from the final release (StatsSA, 2008*a*).[8]

[5] OHS data exist for 1993, but we exclude this year because the 1993 survey had a different sampling frame to the later surveys in that it excluded the 'independent' homelands of Transkei, Bophuthatswana, Venda, and, Ciskei.

[6] The 2004 Master Sample used for GHS 2005-7 was stratified at the district council level, although StatsSA caution that the data is not representative at this level and more recently released versions of the GHS for 2002-2007 do not include a district council variable (DataFirst, 2015).

[7] For the time period we ultimately settle on for the remainder of the thesis, 1995-2011, we have 15 surveys including about 1.5 million people and 385 thousand households.

[8] StatsSA (2010*b*) also point out that researchers should keep in mind for reasons of comparability with previous surveys that from 2009, sampling happened over three months, instead of only in July as was previously the norm for the GHS.

3.3.2 Weight calibration of StatsSA household survey data

As mentioned initially, household surveys only sample a portion of the population and therefore need to be weighted so that sample estimates resemble true population estimates. Weighting procedure is informed by sampling procedure. In the OHS and the GHS, sampling practice is that should a household be sampled, everyone living in that household is sampled. This means that the chance an individual is sampled is equal both to the chance their household is sampled, as well as, the chance their other household members will be sampled. The implication is that weights for members of the same household should be equal and that this weight should also equal the household weight. In cases where household and person data is released in separate files, like the OHS and the GHS, the weight for a given household in the household file should equal the weight for a member of that household in the person file. This is rarely the case in the OHS or the GHS.

For the OHS, StatsSA calibrated the person weights in such a way that people within a household had different weights, i.e. weights were not integrated (Branson and Wittenberg, 2014). This raised the problem of what the household weight should be and resulted in the release of a separate household weight (and for most years of the OHS it is unclear how the final household weight was calibrated[9]). Having two different and unrelated person and household weights means that in some cases researchers will arrive at two different estimates of the same statistic, even though they are using the same set of data. An example of this outcome can be viewed in the discrepancy between the OHS-GHS household-weighted (HHW) and person-weighted (PW) series in both panels of Figure 3.1 in the OHS period, 1995-1999.

From 2002 and with the beginning of the GHS series, StatsSA used integrated weights. That is, they ensured that people within the same household had the same weight; and then used this as the household weight. Although this was consistent with sampling practice, concerns were raised when these weights yielded unrealistic trends in the number of households. An uneven trend came out over the period 2002-2008 (StatsSA, 2010*c*), observable in the Old GHS series plotted in Panel (a) of Figure 3.1. This was unsatisfactory because an accurate series of household counts should be an immediate output of national household survey data. Furthermore, pressure on this point was applied at the time, in particular, by policymakers who required accurate household counts for planning purposes. The uneven trend was possibly a consequence of the weights being calibrated using only person, and not household, auxiliary information.

Capacity and time constraints within StatsSA in 2009 meant that to overcome the technicalities of simultaneously squaring a person and household population, the household and person weight calibration was instead separated from 2009 to present. The person weights continue to be integrated weights, equal within the household and calibrated only on person information. The household weight, though, is now calibrated completely independently using what we call the StatsSA Headship Model that uses only household-level information (StatsSA, 2010*c*). StatsSA then released reweighted versions of the 2002-2008 GHSs with this new weighting procedure (StatsSA, 2010*c*); these reweighted household weights were then the ones to end up in the OHS-GHS series (and are plotted as the 'GHS HHW' in Figure 3.1).

Like in the OHS, having different person and household weights leads to multiple inconsistent estimates. The problem in the case of the GHS is even more pronounced as the two weights are completely

[9]For example, metadata from years 1997-9 describes designing the household weight, but only post-stratifying the person weight (StatsSA, 1997, 1998, 1999). In 1994, though, the weight of the household head was used as the household weight (Wittenberg, 2008).

divorced from each other due to their calibration models being based on different input information. The inconsistency this introduces into statistic estimation is again evident from the disagreement between the OHS-GHS (HHW) series and the OHS-GHS (PW) series in both panels of Figure 3.1 in the GHS period, 2002-2015. In Panel (b), the person weight yields a smooth, evenly-increasing population series, but the household weight produces a highly divergent and unstable series for the same statistic. The opposite is true for the case of households in Panel (a). This is evidence that the person and household universe are not reliably represented concurrently using either weight. Researchers must choose between 'two worlds': one in which people are consistently represented, but not the households they live in, or, vice versa.

3.3.3 Implications of the StatsSA dual-weight system

The StatsSA dual-weight system leads to two main problems: inconsistency and incoherence. These two problems restrict the number of research questions the data can be used to answer, undermining the investment of public funds in their collection. The first problem of inconsistency has already been demonstrated by the disagreement between the differently weighted series in Figure 3.1. Multiple weights introduce noise into inference because this allows multiple estimates per statistic. Being able to extract two different total household counts for a given year in the period is a serious concern for this thesis, in particular.

Estimation is further complicated by uncertainty surrounding how to weight combinations of person- and household-level information. For example, which weight should be applied when counting the number of *people* living in *households* with access to electricity? If researchers want to weight this estimate, they are forced to confront the absurd-sounding choice of, is this more of a person outcome or a household outcome? This question makes no sense because these outcomes cannot be neatly divided into person-only and household-only; they relate to each other. This illustrates that as soon as researchers are interested in questions that intersect the person- and household-boundary, conceptual issues start arising. As we explained in Chapter 1, these represent some of the most important questions for tracking the progress of the post-apartheid project. For example, welfare analysis using per capita household income uses a household variable to draw a person-level conclusion. This is also the case for this thesis, which is ultimately interested in change in the way that *people* form *households*. It is especially unclear which weight should be used to weight a household headship rate, when heads are people but their count must match the number of households.

In other words, the problem with the dual-weight system goes beyond estimation. By only allowing either people or households to be accurately represented at a time, any analysis that crosses the person-household boundary is rendered incoherent in a conceptual sense. By separating person and household weights so categorically, this dual-weight system sets up a conceptual framework of reality in which people and households exist independently and in isolation of one another. It implies that person behaviour plays out independent of our groupings into households; households are implied to have behaviour patterns of their own, divorced from the individuals that comprise them who instead are treated as behaving as a single homogeneous unit. This is intuitively and logically at odds with how social scientists think about how people and households relate.

In these surveys, representativeness has been compartmentalised into either a person universe or a household universe because the weighting system precludes a universe where both people and households

are simultaneously represented in a way that reflects true population estimates. The implication is that researchers can only perform either person-level or household-level analysis in isolation, or in-sample analysis if they want to be conceptually coherent. This limitation seriously reduces the number of questions these data can be used to answer. This problem is not restricted to the OHS and GHS, but is present in all StatsSA household and labour market survey data. The Income and Expenditure Survey, the Living Conditions Survey, and Community Survey all collect both household and person information and have separate person and household weights. For the labour market data, the LFS, QLFS, and Labour Market Dynamics Survey, StatsSA only provide a person weight. Any household-level analysis carried out in the labour market surveys is questionable since person weights are not calibrated using household-level information.

Fortunately, calibration can be remedied more easily than other influences on data quality because it happens after the data has been collected, meaning the problems created by the dual-weight system are not incontrovertible. Our goal in this chapter is to test the feasibility of combining the information StatsSA uses in two independent processes, into one procedure. It is not obvious from the outset that this will work, mainly because it is unknown whether the calibration procedure will compile when using both person and household information. As we explained above, calibration procedures can fail to compile if too many constraints are imposed on the data; or, if the constraints are inconsistent enough with each other, that the calibration cannot satisfy them all at the same time. That is, if the requirements imposed by the constraints are so different that certain cells are pulled in too many different directions. Since StatsSA household survey data has not been calibrated before using total person and household counts simultaneously, these are both potential pitfalls.

Should we succeed, we also need to consider that accurate household counts not forthcoming from the original GHS weights were the impetus behind the creation of the dual-weight system in the first place. This means, a series of accurate household counts is a key outcome we need to achieve, both as the mark of good quality survey weights and for the purposes of this thesis in particular. Combining household and person information into one weighting procedure will restore the concurrent representativeness of people and households in the data, and bring the weighting in line with international best practise (Särndal, 2007). In the next section we put forward our alternative calibration technique, cross-entropy estimation, and describe how we apply it in the South African case.

3.4 Recalibrating survey weights using cross-entropy estimation

Our goal is to reweight the full data series of the OHS and GHS spanning 1994-2015 using cross-entropy estimation in order to achieve more consistent and conceptually coherent person and household counts over the post-apartheid period. The aim is to achieve a single consolidated weight that assigns the same weight to everybody living in the same household, as well as, to the household, itself. This then clearly and logically links people to the households they live in, expanding the conceptual framework in a sensible way, and restoring representativeness to the data. An immediate deliverable is an integrated series of cross-entropy weights representative of both people and households at the same time, that will allow researchers to extract accurate person and household counts. This weight will consolidate all the same information StatsSA uses to calibrate their separate weights into a single model.

The generalities of the calibration problem are essentially the same as those laid out in Section 1, with the cross-entropy technique using an entropy measure as its distance function. Given moment constraints and a normalisation restriction, cross-entropy minimises the information ($\mathbf{I}(\mathbf{p}, \mathbf{q})$) needed to make the new distribution of weights ($\mathbf{p}$) resemble as much as possible information we already have about what it should look like, which could be a distribution of existing StatsSA calibrated or design weights ($\mathbf{q}$). This can be formalised as:

$$\mathbf{I}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} p_i \ln(\frac{p_i}{q_i}) \tag{3.3}$$

where $\mathbf{I}(\mathbf{p}, \mathbf{q})$ is minimised subject to:

$$y_j = \sum_{i=1}^{n} X_{ji} p_i, j = 1, ..., J \tag{3.4}$$

$$\sum_{i=1}^{n} p_i = 1 \tag{3.5}$$

where there are J population moments; y_i is the population mean of the random variable X_j; and equation 3.5 is the normalisation restriction. This constrained optimisation problem can be solved by maximising the unconstrained dual cross-entropy objective function:

$$L(\lambda) = \sum_{j=1}^{J} \lambda_j y_j - ln[\Omega(\lambda)] = \mathbf{M}(\lambda) \tag{3.6}$$

where $\Omega(\lambda)$ is given by:

$$\Omega(\tilde{\lambda}) = \sum_{i=1}^{n} q_i \exp(\mathbf{x}_i \tilde{\lambda}) \tag{3.7}$$

Golan et al. (1997) show that this function behaves much like maximum likelihood. The function $\mathbf{M}$ can be characterised as an expected log likelihood where $\mathbf{p}(\lambda)$ is the exponent and the parameter is λ. The parameter λ reveals the extent to which the new distribution, $\mathbf{p}$, is a distortion of the original underlying distribution, $\mathbf{q}$. Each constraint in the constraint matrix has an associated λ coefficient indicative of how informative that particular constraint is in estimating $\mathbf{p}$. A higher λ coefficient, for example, is an indication that the respective constraint resulted in the $\mathbf{p}$ distribution moving relatively further away from the original $\mathbf{q}$ distribution.

In other words, the technique weights the sample according to information about what the sample looks like; was designed to look like; as well as, information about what the actual population looks like. We use the *maxentropy* command from Wittenberg (2010) in Stata 15 to practically carry out the estimation. The distribution of prior information, $\mathbf{q}$, comes from the design weights in the case of the GHS, and the calibrated weights in the case of the OHS. Population moments enter the command via a constraint matrix which is set up based on auxiliary information from StatsSA.

The sample for each year is obtained from the publicly accessible data set downloaded from South African data repository DataFirst's website (StatsSA, 2010-2013, 2011-2018*b*). Design weights are not publicly released for either the OHS or the GHS. The design weights for the GHS period 2002-2011 were received in private correspondence between Martin Wittenberg and StatsSA. As a result of only having

weights for this period, we only run the recalibration up until 2011. We have to use StatsSA's calibrated weights in the case of the OHS, and these were available in the publicly available data sets.[10]

The constraint matrix consists of 103 constraints, which are all proportions of the total person population: 63 individual age-sex-race categories; 8 province categories; and, 32 age-sex-race household headship categories.[11] The individual-level and province population moments come from StatsSA (2018*a*,*c*) MYPE, which are publicly available for the period 2002 to present on the StatsSA website (i.e. the GHS period). Population estimates for the OHS period came from back projections of the population from DataFirst (2018). This was necessary as StatsSA do not provide disaggregated-enough population estimates for our purposes for these years.[12]

The household moments come from household headship rates calculated by StatsSA for their household weight model, the StatsSA Headship Model, also received by private correspondence between StatsSA and Martin Wittenberg. StatsSA calculate these headship rates using the ten percent samples of the three censuses (1996, 2001 and 2011), as well as, the 2007 Community Survey. Using the StatsSA numbers for 1996, 2001, 2007 and 2011, we then interpolate headship rates for intervening years. Effectively, constraints enter as shares the new weights must reflect when adding up to the overall person total. The total household count is then inserted into the constraint matrix via the benchmarked share of the total population who are household heads.

The constraints are set up in the person-level data file for a particular year, e.g GHS 2006, by creating dummies that correspond to the categories in the constraint matrix. The person-level file is then collapsed into a household-level file converting person-level constraints into proportional household-level constraints; but, summing the person weights within households so the file can reflect the population total, even though it is household-level. The *maxentropy* command is then run on this household file and generates a new set of weights subject to the constraint matrix, the prior weights, and the condition that the new weights add up to the person population total (from the MYPE).

The outcome is a series of household weights incorporating both person and household information in a household file that sums to the person population total. We divide this weight by the household size of each household to reach an integrated person-level weight. These person-level weights are then merged back into the person-level file and reattached to individuals. This process compiled successfully for all years in the interval 1994 - 2011. This series of cross-entropy weights is then incorporated into the OHS-GHS series. There are now three weights in the OHS-GHS series: the StatsSA person weight; the StatsSA household weight; and the new cross-entropy weight (CEW).

[10] Although we plan to continue investigating, it appears there are no design weights still on record for the OHS at StatsSA.

[11] The 64th individual and 9th province category are excluded as these sets are mutually exclusive. There are eight age categories for individuals in intervals of ten years, beginning with 0 - 9 years and ending with those aged 80 years and above. There were four age categories for the household heads. These were 0 - 34 years; 35 - 49 years; 50 - 64 years; and 65 years and older.

[12] The population was back-projected by applying an exponential model of population growth to StatsSA MYPE from 2002 onwards and using population growth rates from the Actuarial Society of South Africa (ASSA).

3.5 Comparing the cross-entropy weight to the dual-weight system

There are two problems with the StatsSA dual-weight system: inconsistency introduced into estimation and an incoherent implicit conceptual framework of people and households. In order to show that the CEW is a useful alternative to the dual-weight system, it needs to perform better on both these counts. Since there is only one weight for people and households, we already know that the cross-entropy weight won't be producing inconsistent estimates; by which we mean, there will only be one estimate per statistic. Instead, what is worth evaluating, is the soundness of the calibration itself. This can be divided into assessing distortion and replication. Respectively, this involves assessing the degree to which the new weight is a distortion of the prior weighting distribution (the **q** distribution) and whether it is able to reproduce its own constraints (the census and MYPE). This is covered in Section 3.5.1. In Section 3.5.2, we evaluate the coherence of the conceptual foundation of each weighting system - StatsSA dual-weight and cross-entropy - by the degree to which each system is internally consistent.

3.5.1 Evaluating the validity of the cross-entropy calibration

Distortion

We begin by comparing summary statistics of the CEW with summary statistics from the prior **q** distribution in Table 3.1. The **q** distribution came from the person weight in the case of the OHS and the design weight in the case of the GHS. Table 3.1 reports the range of the two weight distributions in each year of the series, as well as some diagnostics about the λ coefficient output which is a signal about the health of the calibration. The λ coefficient is a measure of the extent to which the prior distribution is distorted by the constraints to yield the CEW distribution. High λ coefficients are therefore indicative of constraints that have resulted in a CEW distribution that is substantially different to the prior weight distribution. There is no particular rule regarding how high a λ is too high, but five is a useful rule of thumb (Wittenberg, 2010), reported in the last column.

The CEW tends to range higher than the prior distribution, especially in the OHS period. The year that stands out the most is 1994 which has a minimum CEW of zero[13] and an enormous maximum of over 200 thousand. The 1994 calibration is the weakest based on the λ coefficient diagnostics: about a third of coefficients are higher in absolute terms than five. It should be noted, though, that the 1994 sample is unusual for being different in important ways to the sample used for the rest of the series. In 1994, whites were oversampled and Black Africans were undersampled. In a sense then, high λ coefficients in 1994 are indicative of constraints that are highly informative not so much in distorting the sample, as 'distorting it back' to what it should look like.

Some of the highest λ coefficients came out for older, usually male, age groups for both the individual and household headship categories. The exception was a high degree of adjustment required for the shares of Black African women over 70 years in the GHS period in particular. In the OHS period, Asian/Indian

[13]The zero CEW weights in 1994, 1997, and 1998 are indicative of missing data required for the calibration. In addition to seven cases of zero-valued CEWs, there are 79 cases of missing CEWs. In the case of the seven zeroes, these occurred in 1997 and 1998 when the PW was missing completely - the weight used as the prior distribution for the CEW. There were three missing CEWs in 1996 and 76 in 1998, all of which coincide with missing PWs and sometimes also household heads.

and Coloured men over 80 years needed more adjustment. The headship constraints in general were more informative than the individual and provincial ones and this was particularly the case for white male heads over 50 years. These patterns are likely reflective of smaller samples of older people, especially males.

The calibration settled down from 1995 onwards. The shares of absolute λ coefficients that exceed five are negligible after 1994 and - aside from 1995 - the average absolute λ is always less than two. The 2005 calibration is noteworthy because it has an especially high maximum, but the rest of the λ coefficient diagnostics do not draw attention. Noting the weakness in 1994, we conclude that the calibration is satisfactorily similar to the prior distribution.

Table 3.1: Summary and diagnostic statistics for the prior weight distribution and the CEW in the OHS-GHS series

	Prior Distribution		**CEW**		**Lamda (λ) Diagnostics**		
Year	**Min.**	**Max.**	**Min.**	**Max.**	**| Mean |**	**| Max. |**	**Share of $\lvert\lambda\rvert \geq 5$**
1994	14.43	1 936.71	0.00	206 570.90	4.10	13.21	0.32
1995	0.07	1 759.65	0.12	6 134.33	2.22	8.48	0.07
1996	61.77	6 053.41	11.74	6 432.31	1.57	6.91	0.03
1997	42.00	1 834.00	0.00	5 506.44	1.87	5.78	0.03
1998	47.66	2 629.73	0.00	7 432.17	1.10	4.82	0.00
1999	12.71	2 387.87	13.13	9 522.72	1.45	5.68	0.03
2000							
2001							
2002	16.38	8 437.75	6.73	9 384.52	1.78	5.85	0.02
2003	6.35	7 986.57	2.50	12 096.30	1.82	6.86	0.02
2004	8.04	10 264.53	4.88	10 962.53	1.37	5.05	0.01
2005	5.90	5 969.12	3.74	20 056.11	1.19	3.95	0.00
2006	5.89	11 435.05	1.45	10 053.11	1.41	4.47	0.00
2007	5.89	7 277.31	1.19	11 472.63	1.12	4.18	0.00
2008	115.43	7 275.32	10.95	9 061.39	0.94	3.89	0.00
2009	115.43	5 171.03	10.84	8 536.67	1.46	5.02	0.01
2010	96.72	5 265.66	20.21	10 194.55	1.12	4.89	0.00
2011	96.72	5 887.85	20.91	9 921.07	1.04	4.60	0.00

Notes: own calculations. Prior Distribution = StatsSA calibrated person weight for the OHS years and design weights for the GHS years. Min. = Minimum; Max. = Maximum; | Mean | = Absolute mean; | Max. | = Absolute maximum; Share of $\lvert\lambda\rvert \geq 5$ = Share of constraints for the respective year with an absolute value of λ exceeding 5.

Replication

Figure 3.2 comprises the main result displaying the quality of the CEW in estimating population and household counts in comparison to the StatsSA person (PW) and household (HHW) weight. For reference, these figures are reported in table format in Table 3.2. All three weights are trying to match the estimates from those of the census or MYPE. The CEW outperforms both the PW and HHW by this criterion. In Panel (b), the CEW line lies exactly on top of the census and MYPE estimates of the population for the full period 1994 - 2011. The PW matches the MYPE exactly in the GHS period, but slightly

undercounts the population in the OHS period compared to the CEW.[14] The HHW, on the other hand, produces a jagged trend line that appears quite unrelated to the census/MYPE population trend. Further to this, Table 3.2 records that the CEW population counts never deviate from the MYPE by more than an order of 50 000 people; and, the difference is always at least the same as, but usually smaller, than the discrepancy between the PW and the MYPE. That the CEW is able to reproduce the census or MYPE so well demonstrates the success of the calibration since these estimates were used as constraints.

Turning to household counts, there is a minor discrepancy between the CEW household count and the census count in Table 3.2. The largest margin of error occurs in 1996: the census counted 9.06 million households in 1996, the CEW overestimates at 9.25 million. For the census 2007 and 2011, the CEW over- and underestimates, respectively, by 100 000 households. This is a consequence of slightly different total population counts entering into the calibration at the same time in census years. The population count enters directly into the calibration as a total count from the MYPE; but, also indirectly via the household headship constraints from the StatsSA Headship model based on the census ten percent samples. These two population counts differ slightly.[15] Since the direct population counts were sourced from official census publications, they are more likely to accurately reflect weighting for non-response. In years 1996 and 2007, the direct population count exceeded the indirect population count, logically leading to an overcount on CEW's part as cross-entropy uses proportional constraints, and vice versa in 2011. Ultimately though, the calibration is reproducing its constraints. The discrepancy from the census in Table 3.2 is not a reflection of a poor quality calibration but of disagreement in the constraints, themselves.

Despite this small dissimilitude, the CEW is able to match the census best over the entire time period in Panel (a) of Figure 3.2 by accommodating the 2001 census estimate much better than the HHW. The HHW fits the census estimate in 1996 better than the CEW, but the CEW has a smoother trend over the OHS period compared to the step-wise pattern of the HHW between 1994 and 1999. Clearly the HHW has been calibrated well in census years, but the quality of the calibration is questionable in non-census years. This serves to underline the role of good quality weighting: in non-census years when population moments are not readily available, we rely upon the soundness of our weighting practise. We can conclude that the calibration of the CEW is sound based on its resemblance to the prior $\mathbf{q}$ distribution and its strong performance in reproducing its constraints, with perhaps the exception of 1994. The CEW yields one estimate per statistic which is ideal, and performs best over the span of the time period when compared to the PW or HHW.

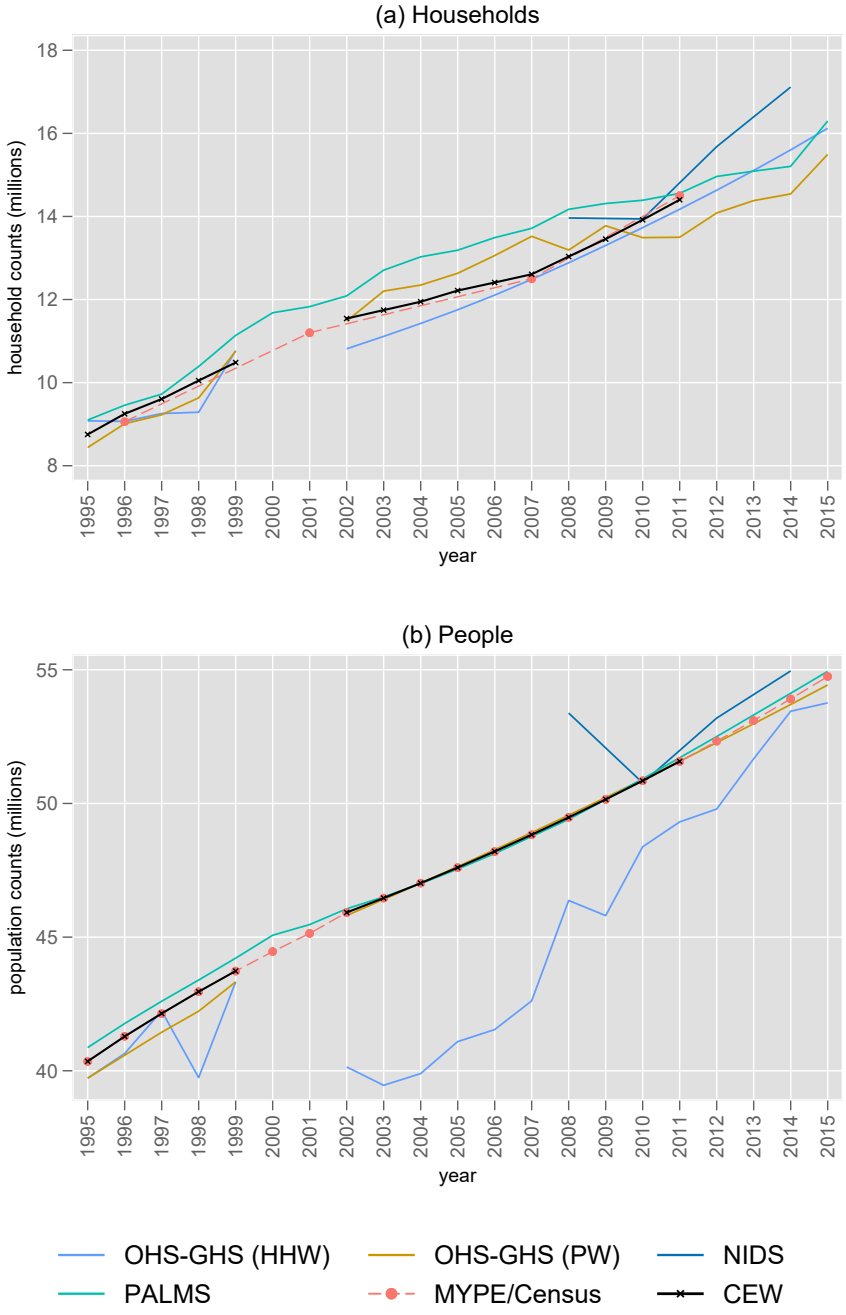
3.5.2 Assessing the internal consistency of the cross-entropy weight and the dual-weight system

Since social scientists think people and households interact in sensible meaningful ways in the real world, a desirable property of the data used to study them is that it should do the same. For our purposes then, internal consistency in household survey data means that the two main units of analysis - people and households - relate to each other in a logical way and yield estimates that do not contradict each

[14] Note that the MYPE for the OHS period do not come from StatsSA directly as detailed in the figure notes.

[15] In 1996, the direct (MYPE) population was 41.3 million and the indirect (StatsSA Headship Model) one was 39.7 million. In 2007, the direct population was 48.8 million and the indirect, 48.5 million. In 2011, the direct population was 51.6 million and the indirect was 51.8 million.

Figure 3.2: Comparing total population and household counts in South Africa to the CEW-weighted OHS-GHS series, 1995-2015



Notes: own calculations using HHW = Household weighted OHS-GHS series; PW = Person weighted OHS-GHS series; CEW = cross-entropy weighted OHS-GHS series; NIDS = National Income Dynamics Study; Old GHS = GHS weighted using the original integrated weight; PALMS = Post-Apartheid Labour Market Series v3.3; Census = household totals from census 1996, 2001, and 2011 and Community Survey in 2007 in Table 3.10 in StatsSA (2012*c*) and Table B in StatsSA (2001*a*); MYPE = mid-year population estimates from StatsSA (2018*a*) for 2002-2015, and DataFirst (2018) for 1995-2001.

Table 3.2: Comparing official population and household counts to estimates from the OHS-GHS series using different weights, 1994-2011

	Household Counts (000 000's)				Population Counts (000 000's)			
Year	**PW**	**HHW**	**CEW**	**Census**	**PW**	**HHW**	**CEW**	**MYPE***
1994	8.89	7.10	8.32	-	40.30	30.30	39.40	39.37
1995	8.93	9.08	8.75	-	39.70	39.70	40.40	40.35
1996	9.09	9.07	9.25	9.06	40.60	40.60	41.30	41.29
1997	9.16	9.26	9.61	-	41.40	42.20	42.10	42.14
1998	9.87	9.29	10.00	-	42.20	39.70	43.00	42.96
1999	10.80	10.80	10.50	-	43.30	43.30	43.70	43.73
2000	-	-	-	-	-	-	-	44.46
2001	-	-	-	11.21	-	-	-	45.14
2002	11.50	10.80	11.50	-	45.80	40.10	45.90	45.92
2003	12.20	11.10	11.70	-	46.40	39.50	46.50	46.46
2004	12.30	11.40	11.80	-	47.00	39.90	47.00	47.02
2005	12.60	11.80	12.20	-	47.60	41.10	47.60	47.60
2006	13.10	12.10	12.40	-	48.30	41.50	48.20	48.20
2007	13.50	12.50	12.60	12.50	48.90	42.60	48.80	48.83
2008	13.20	12.90	13.00	-	49.60	46.40	49.50	49.48
2009	13.80	13.30	13.50	-	50.20	45.80	50.20	50.15
2010	13.50	13.70	13.90	-	50.90	48.40	50.90	50.85
2011	13.50	14.20	14.40	14.50	51.60	49.30	51.60	51.57
Change 1996 - 2011	4.41	5.13	5.15	5.44	11.00	8.70	10.30	10.28

Notes: own calculations using the OHS-GHS series. HHW = Household weighted; PW = Person weighted; CEW = Cross-entropy weighted; Census = household totals from census 1996, 2001, and 2011 and Community Survey in 2007 in Table 3.10 in StatsSA (2012*c*) and Table B in StatsSA (2001*a*); MYPE = mid-year population estimates from StatsSA (2018*a*); * MYPE for the period 1994 - 2001 are provided by DataFirst (2018) back projections.

other. In order to investigate internal consistency, we use the statistic of average household size, following similar work by Branson and Wittenberg (2014). Panel (a) of Figure 3.3 presents average household size in South Africa for the period 1995 - 2015. The CEW yields a smoother more realistic trend than the jagged trend estimated by both the HHW and the PW. Most importantly, the CEW fits the census estimates of household size quite precisely, unlike the StatsSA weights. This is already a clue that the CEW is more internally consistent than the HHW and PW.

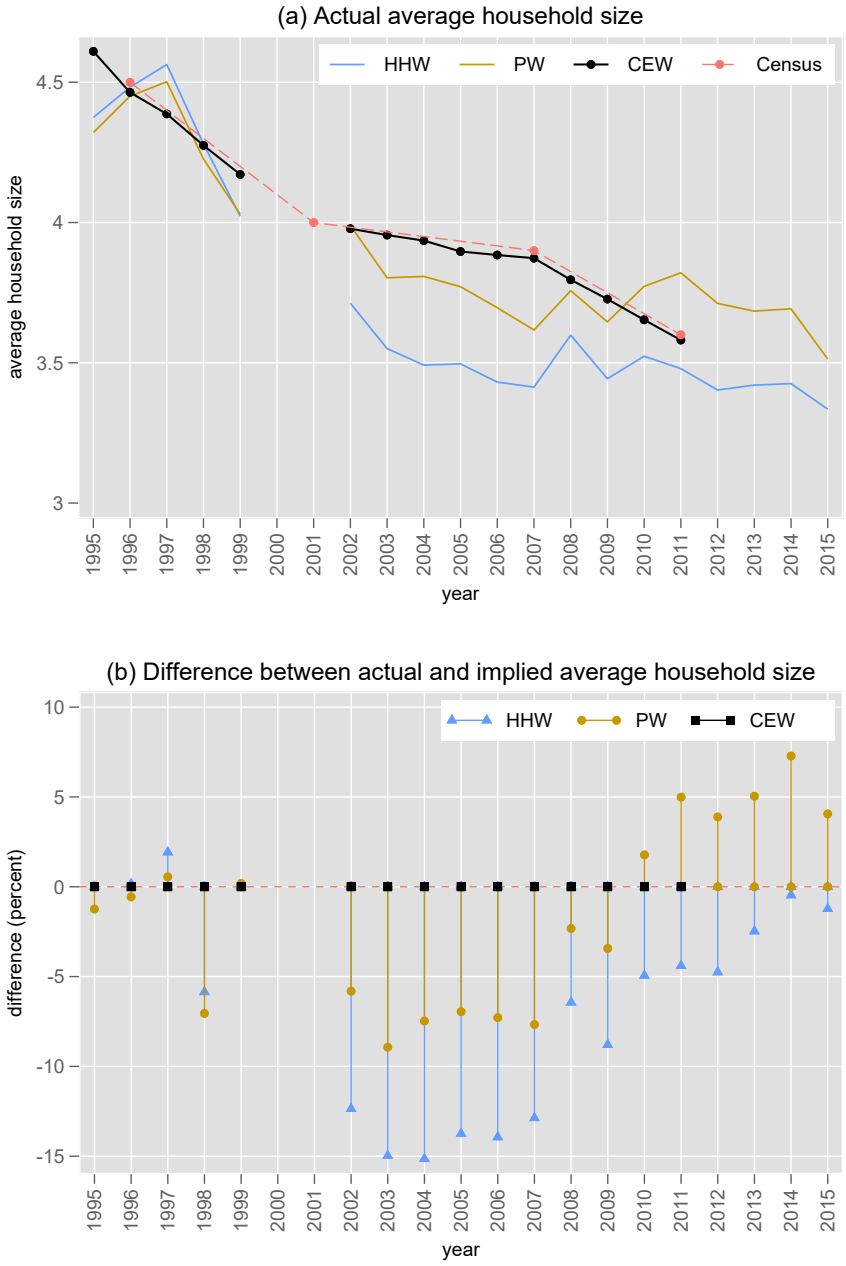
Average household size can be used to investigate internal consistency for two reasons: it is calculated using both person and household information; and, it can be calculated in two ways allowing for inspection of contradictions. Actual average household size is the expected value of a derived household size variable in the data set taken over the distribution of households. Implied average household size, on the other hand, is determined by dividing the total population by the total household count:

$$Actual = \mathbf{E}_H[hhsize_h] \tag{3.8}$$

$$Implied = \frac{N}{H} \tag{3.9}$$

where $hhsize$ is the number of people living in household h, N is the total population, H is the total household count.

Figure 3.3: Testing internal consistency using average household size in the OHS-GHS series, 1995-2015



Notes: own calculations. HHW = Household weighted OHS-GHS series; PW = Person weighted OHS-GHS series; CEW = Cross-entropy weighted OHS-GHS series; Census = census ten percent samples and Community Survey in 2007. Actual household size = mean of household size variable in the data set and weighted as specified; Implied household size for HHW and PW = total population (person weighted) divided by the total household count (household weighted); Implied household size for the CEW = total population count (CEW weighted) divided by total household count (CEW weighted).

In the case of the StatsSA weights, implied household size is calculated by dividing the *person*-weighted population by the *household*-weighted household counts. If the weighting system is internally consistent, these two methods should produce equal estimates of average household size; that is, actual should equal implied. This is evidently the case for the CEW in Panel (b) of Figure 3.3. This figure reports the percentage difference between actual and implied average household size. The CEW has a difference of zero for all years. By contrast, there is frequently a large discrepancy between the actual and implied average household size using the HHW and the PW.

Based on the above sub-section and this one, we conclude that the CEW presents a useful alternative to the existing StatsSA dual-weight system of the PW and the HHW. The previous sub-section demonstrated the soundness of the cross-entropy recalibration in general and also showed that the CEW performs better than the HHW and PW in estimating population and household counts over the period of time it covers. This sub-section argued that conceptually the CEW is superior since it is internally consistent in a way that the dual-weight system is not. Taken together, this means the CEW solves the two problems associated with the dual-weight system by providing singular consistent estimates and being conceptually coherent. In this way, we are able to restore representativness to the OHS-GHS series as the CEW allows the data to consistently represent both people and households at the same time.

3.5.3 The performance of the cross-entropy weight on non-constraining statistics

We are confident that the CEW presents a valid alternative to the StatsSA dual-weight system. However, there are limits on the extent to which reweighting can improve the quality of the data. Earlier it was explained that the degree to which a sample is representative depends on factors at all phases of the data collection, processing, and, dissemination process; post-sampling survey weight calibration being just one of these phases. Representativeness can also be affected by sample design, non-response, measurement error, enumerator practice, and luck surrounding the drawing of a single random sample. Weight calibration happens after all these factors have already been set, meaning that if there are serious deficiencies with the underlying sample, reweighting is a relatively limited tool to correct representativeness.

Limitations on the power of the reweighting mean that whilst the CEW might perform better than the dual-weight system on certain statistics, this might not be the case for all statistics. Results so far have focused on population and household counts, statistics we specifically used to constrain the calibration. In other words, population and household counts are statistics for which we would expect the CEW to perform well, because it was specifically designed to do so. Another good sensitivity test of the quality of the two weighting systems is to use statistics with which the weights have not been specifically constrained.

In Figure 3.4, we compare the performance of the different weights to the census for four different statistics, chosen to test different aspects of the weights. The first of these in Panel (a) is the share of single-person households. This is a statistic about which researchers have long raised concerns about measurement accuracy (Branson and Wittenberg, 2014; Kerr and Wittenberg, 2015; Wittenberg and Collinson, 2007). Wittenberg and Collinson (2007) note that the proportion of single-person households increases far more rapidly in national surveys than it does in the location of Agincourt, Mpumalanga, where the Agincourt Health and Demographic Surveillance System has been collecting longitudinal data

for many years. Kerr and Wittenberg (2015) explain that the share of single-person households are underestimated in the OHS because of undersampling of small households and hostel-dwellers. These authors describe the abrupt increase in single-person households between the end of the OHS series and onset of the LFS series as "improbable". Such considerations suggest that there may be problems with the underlying sub-sample of single-person households, undermining the power of reweighting to fix this series.

Panel (a) in Figure 3.4 confirms these concerns. None of the weighted series connect with the census data points in a precise way. This is particularly prominent at the beginning and end of the period. All three of the weights underestimate the share of single-person households in 1996 by more than 6 percentage points, and by 5 percentage points or more in 2011. Imprecision on the part of all three weights is likely driven by the undersampling of single-person households which appears to be happening throughout the OHS-GHS series. Calibration method also appears to be playing a role; although all weights use the 2007 Community Survey as a benchmark, this benchmark appears more restrictive for the CEW. We revisit this problem in detail in the next chapter because, as we explained in the introduction to this thesis, single-person households are an important constituency when studying household formation.

Panel (b) of Figure 3.4 plots another household-level statistic, which is the share of households that are urban. There are notable discontinuities in the trends which coincide with change of Master Sample. As we described in Section 3.3, years 1996-2004 all use a Master Sample based on the 1996 census, although the sampling did change slightly when the GHS started. With the splitting of the OHS into the LFS and GHS, the Master Sample was split into five clusters with the LFS drawing from two clusters (it ran twice a year) and the GHS from a different cluster. GHS 2005-7 are based on the 2004 Master Sample; GHS 2008-2014 are based on the QLFS Master Sample; and GHS 2015 is based on the 2013 Master Sample. The CEW does not perform very differently to the HHW or PW and equally reflects the discontinuity arising from change in Master Sample. This figure is another good demonstration that the reweighting cannot fix underlying sampling. Potentially, one way to get around this problem might be to constrain the calibration on urban location.

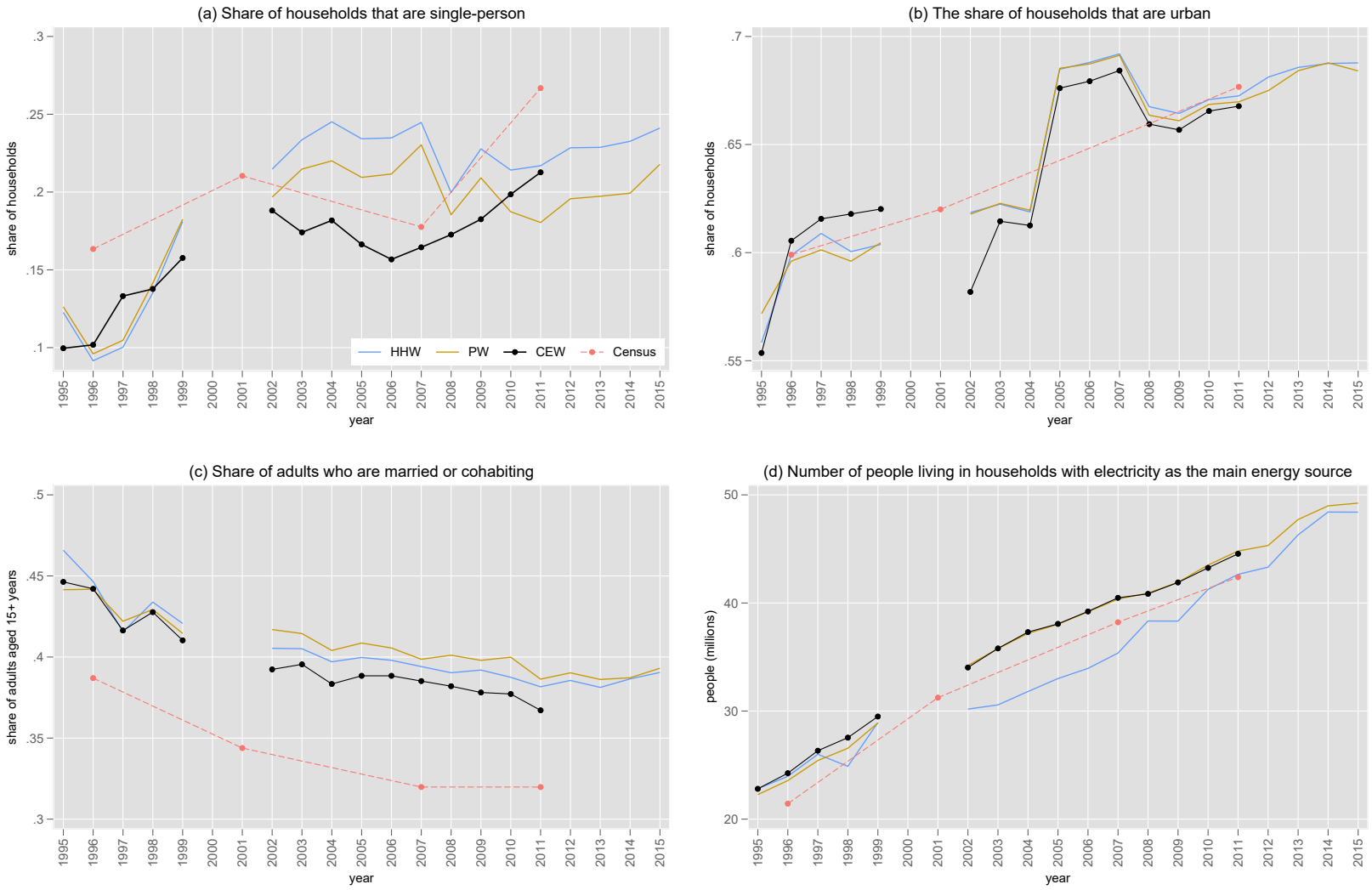
In Panel (c), we turn to a person-level statistic, which is the share of married or cohabiting adults. There is a disconcerting level of overestimation of this share by all the survey estimates compared to the census which is potentially a function of survey practise. It's quite possible that married people are more likely to be enumerated in general, than their not-married counterparts; for example, because field workers may be more likely to find one member of the couple at home. Of the set of weighted series, though, the CEW is the least biased. The CEW series lies below both the HHW and PW series demonstrating that just person (PW) or household (HHW) information alone does not yield as accurate a series compared to if both types of information are used. It is notable that the PW series is furthest from the census even though this is a person-level statistic and the PW is the person weight.

Finally, in Panel (d), we turn to a person-level statistic based on a question from the household file: the number of *people* living in *households* with electricity. This connection between the person- and household-level is one the CEW could be expected to be better at than the StatsSA weights. However, Panel (d) shows that the CEW performs as well as the PW in this case, and is very slightly further away from the census than the PW in the OHS period. The CEW does not appear to constitute an improvement in Panel (d), but it also doesn't appear to be doing any harm.

The purpose of these figures is to make the point that reweighting is not a panacea for inadequate

representativeness. The CEW will not necessarily outperform the dual-weight system on every statistic (relative to the census). Underlying problems with the sample - such as seriously undersampling a certain sub-sample - are not something that can be easily remedied with reweighting. What the CEW offers is not a correction to every series of statistics in the OHS-GHS series, but rather a sturdier foundational framework on which later analysis can be more reliably built. The CEW allows the series to have two dependable cornerstones, people and households, instead of one, and this substantially improves coherence of the base on which later analysis is built.

Figure 3.4: Comparing the performance of various weights in the OHS-GHS series to the census



Notes: own calculations. HHW = Household weighted OHS-GHS series; PW = Person weighted OHS-GHS series; CEW = Cross-entropy weighted OHS-GHS series; Census = census ten percent samples and Community Survey in 2007, weighted using the household weight in panels (a) and (b), and person weight in panels (c) and (d).

3.6 Conclusion

To study household formation in this thesis, we have stacked a series of StatsSA cross-sectional household surveys to construct the OHS-GHS series. Before we can use this series, we had to deal with a key data quality issue embedded in the source data, the survey weights. In the versions of StatsSA household survey data that are currently publicly released, household weights and person weights are calibrated in totally independent processes resulting in weights that are never equal to each other, even though sampling practise implies they should be. This creates two main problems: noise is introduced into estimation; and, an incoherent conceptual framework is cemented into weighted analysis using both household and person information. This frustrates the ability to answer many research questions, but the questions for this thesis in particular. It undermines our ability to extract accurate household and person counts for the purposes of plotting trends. And, it undermines our ability to coherently investigate how *people* are forming *households*. For example, it is far from clear how to weight a household headship rate when heads are people, but their count needs to tally with the number of households.

To solve these problems, we successfully managed to create a new series of cross-entropy weights that align with the survey sampling practise; yield household and person counts that conform with census benchmarks; and which reconcile people with the households in which the live. In this manner, we have restored the mutual representativeness of people and households to the data, and thereby, the number of research questions these data can be used to answer. Since our new weights are only available until 2011 and seemed unreliable for 1994, we limit our period of study to 1995-2011 for the remainder of the thesis.

It is worth stressing that reweighting is not a panacea for every data quality issue. There are limitations on what reweighting can solve given the quality of the raw sample, which is influenced in sometimes permanent ways at numerous points in the data collection process. Slightly tuning the weights will not compensate for fundamental sampling problems, like non-coverage, for example. The purpose of the reweighting is not a wholesale improvement of every data output, but rather, a strengthening of the the statistical and conceptual foundation on which later analysis is built. With these caveats in mind, we conclude that, overall, the CEW weights improve the quality of the StatsSA survey data and the degree to which it is fit for the purpose of advancing the post-apartheid research agenda and studying household formation.

That said, this chapter constitutes only the first step in the work to reweight the OHS-GHS data, which continues in the next chapter. This chapter was specifically concerned with whether we could combine the information StatsSA uses in two separate calibration processes, into one procedure. Although this is done for household surveys elsewhere in the world (Särndal, 2007), as far as we know, StatsSA household survey weights have never been simultaneously benchmarked to total person and household counts and it was not obvious at the outset that it would work. Testing whether this could be done at all was therefore the first step towards reweighting the OHS-GHS series. Now that we have shown we can achieve this, we turn our attention to other aspects that might improve the calibration.

Chapter 4

Accounting for small households: reweighting South African household survey data (again)

4.1 Introduction

In this chapter, we build on the work of the previous chapter to construct a series of survey weights to use with the OHS-GHS series for the analysis in Parts II-IV of this thesis. In Chapter 3, the goal was to find out if we could overcome the problems stemming from StatsSA's dual-weight system by integrating the information StatsSA uses in two separate weight calibration models, into one model. We were successfully able to do so, and this process yielded a new cross-entropy weight benchmarked to both total person and household counts. In this chapter, that objective is broadened to also account for other aspects of the calibration model that might be undermining the quality of the data for our purposes; mainly, but not limited to, the undersampling of small households.

Small households are a particularly important constituency for the study of household formation for a number of reasons. Firstly, the share of households that are single-person has risen to unprecedented levels around the world, representing a major demographic trend of the modern age (Snell, 2017). Secondly, in South Africa, single-person households are historically important because of the association between migrant labourers and worker hostels in which the majority of households are often single-person (Xulu, 2014; Machemedze et al., 2007). Thirdly, single- and two-person households are common household structures in which young people might find themselves on their trajectories out of the parental home, at least according to a nuclear family pattern (Ziehl, 2002).

Accurate measurement of small households is, therefore, critical to correctly characterising household formation in South Africa. However, our ability to do so is undermined in two ways in the OHS-GHS series, even with our new cross-entropy weight from Chapter 3: benchmarks and undersampling. The

household benchmarks in the StatsSA Headship Model[1] come from the ten percent household samples of the 1996, 2001, and, 2011 censuses; as well as, the 2007 Community Survey. However, two of these sources are probably providing benchmarks that are too low, for different reasons. The first of these is the 2007 Community Survey, which is a larger-than-usual household survey conducted by StatsSA in lieu of a census (StatsSA, 2007*b*); but which yields estimates that are clearly out-of-step with the actual censuses. On the other hand, the 1996 ten percent household sample excluded worker hostels (StatsSA, 1996), unlike the samples from the later two censuses, making it especially prone to underrepresent small households.

The second source of bias for counting small households is that such households appear to have been undersampled in many years of the OHS-GHS series. Kerr and Wittenberg (2015) have already documented the undersampling of small households in the OHS's, and we present evidence in this chapter that this is continuing in the GHSs, despite improvements to fieldworker practise following the experience in the early OHS's. Additionally, the weights StatsSA release with the data do not compensate for the undersampling, resulting in a 25% undercount of single-person households by 2011 compared to the census estimate, and a trend that moves incongruously with the census series.

In this chapter, we create three more series of cross-entropy weights in addition to the one created in the previous chapter, in which we adjust the constraints to reflect these different issues one at a time. We first remove the 2007 Community Survey which makes a substantial improvement to the health of trends for many outcomes. Then, we inflate the census 1996 benchmark to account for worker hostels. This makes a noticeable difference to the trends in Black African male household formation in the 1990s. Since none of these changes improve estimates of the number and shares of small households, our final weight is specifically constrained on the number of one- and two-person households. The outcome is our 'Best CEW' series of weights, which combines all the advantages of the weight from the previous chapter with more careful attention to benchmarks, and which yields totals of single- and two-person households that agree with the censuses.

The next section elaborates on the importance of small households for the study of household formation. Section 4.3 describes the sampling and benchmarking issues surrounding the measurement of small households in the OHS-GHS series. Section 4.4 describes the weights we calibrate to cope with these different issues using the same cross-entropy method as in Chapter 3. As such, we do not repeat a description of the method, only the updates to the constraints and the motivation behind each. The quality of each calibration procedure as described by the lambda parameter output is summarised in Section 4.5. Section 4.6 is our first set of results comparing the weighted series on key demographic outcomes: household and person counts, average household size, as well as, small households. In Section 4.7, we compare the weighting series for the outcome variables we use in Part II and III of this thesis: household headship and home-leaving. Levels and trends in these variables are more reliable with our Best CEW series. Section 7.7 concludes.

[1]This is the model StatsSA uses to calibrate the household weights in the GHS. We introduced and described this model in the previous chapter.

4.2 Why accurate measurement of small households matters for research on household formation

As we detailed in the literature review in Chapter 2, single-person households have become an important demographic phenomenon on a global scale. In some Western European countries, the share of households that are single-person has exceeded 40% (Ortiz-Ospina, 2019) after being almost unheard of in the first half of the 20th century (Snell, 2017). In developing countries, living alone has slowly started to tick up in Latin America and Africa but is nowhere near the levels of the developed world (Esteve et al., 2020). Southern Africa, though, is an outlier in this regard due to its long history of oscillating labour migration (Wilson, 2001). Southern African countries with a history of sending migrants to South African mines, like South Africa, Botswana, Namibia and Lesotho, stand out for having proportions of single-person households exceeding those of its middle-income peers and on par with other developed countries (Esteve et al., 2020; Ortiz-Ospina, 2019).

In the 1980 census of South Africa, which excluded the independent homelands of Bophuthatswana, Transkei, and Venda, 7% of rural and 6% of urban households were single-person (Simkins (1986) in Ziehl (2002)).[2] In the post-apartheid period, censuses include these territories and the share of single-person households in South Africa grew from about 16% of households in the 1996 census to 27% in the 2011 census (own calculations). This was the same share of single-person households as the United States in 2011 (27%) and more than in New Zealand and Portugal in the same year. The share of single-person households in Brazil and Colombia - countries to which South Africa is commonly compared - came in lower at 12% in 2010 and 11% in 2015, respectively (Ortiz-Ospina, 2019). Questions remain about to what degree Western drivers of living alone apply to South Africa's context, and integral to finding such answers is an accurate portrayal of the country's trend in solo living.

South Africa's history of labour migration also intersects with small households because of the role played by worker hostels (Xulu, 2014; Thurman, 1997). During apartheid, labour migrants employed in urban areas or by mining companies often lived in single-sex worker hostels built by the provincial or local authorities or owned by private companies. The building of these hostels proliferated in the 1960s and 70s and reinforced already-existing patterns of oscillating labour migration. In many instances, wives and families were prohibited from co-residing and only short-term visits controlled by a permit-system were allowed (Thurman, 1997). After apartheid, these restrictions were lifted and women and children moved in, but many hostels fell into disrepair, lacked municipal servicing and reports of violence escalated in correspondence with rising unemployment (Xulu, 2014; Marais and Venter, 2006). Today, the structures of hostels remain and the post-apartheid government has put forward changing policies about how to deal with them; but, in general, they occupy a contested and unresolved space in South Africa's housing landscape (Xulu, 2014; Marais and Venter, 2006).

In the 2001 census, about 3.7% of all households were resident in worker hostels, home to about 640 000 people (own calculations); 43% of people living in worker hostels were living in single-person households (Machemedze et al., 2007), representing 13% of all single-person households (own calculations). In the 2011 census, this was 2% of households home to about 680 000 people, 50% of whom were living in single-person households, but this time only representing just under 4% of single-person households (own calculations). In other words, worker hostels are particularly important for capturing small households,

[2] As Ziehl (2002) points out, the proportions of single-person households in the independent homelands could be expected to be even lower than in 'white' South Africa.

and especially so closer to the beginning of the post-apartheid period.

The historical importance of single-person households in South Africa, combined with the rise of solo-living as a global phenomenon, means accurate counts and trends for small households are important outcomes in their own right. In addition to these reasons, solo-living amongst young people is often a key link in the trajectory out of the parental home, usually associated with men following a nuclear family pattern (Ziehl, 2002), making trends in small households salient to our analysis of home-leaving in Chapter 7.

4.3 Small households in the OHS and GHS: sampling and benchmarking issues

To illustrate the issues surrounding the measurement of small households in the OHS-GHS series, we chart trends in one- and two-person households in Figure 4.1. We focus on one- and two-person households since findings from a study by Machemedze et al. (2007) suggest these were the small household types most seriously affected by the undersampling in the OHS period.[3] Figure 4.1 plots the rate at which these two types of households occur in the unweighted household sample in the 'sample' line; and, the weighted share according to the StatsSA household weight with the 'HHW' line. Change in Master Sample is indicated by red vertical lines. Details about Master Sample shifts were described when we introduced the data in Section 3.3 in Chapter 3.

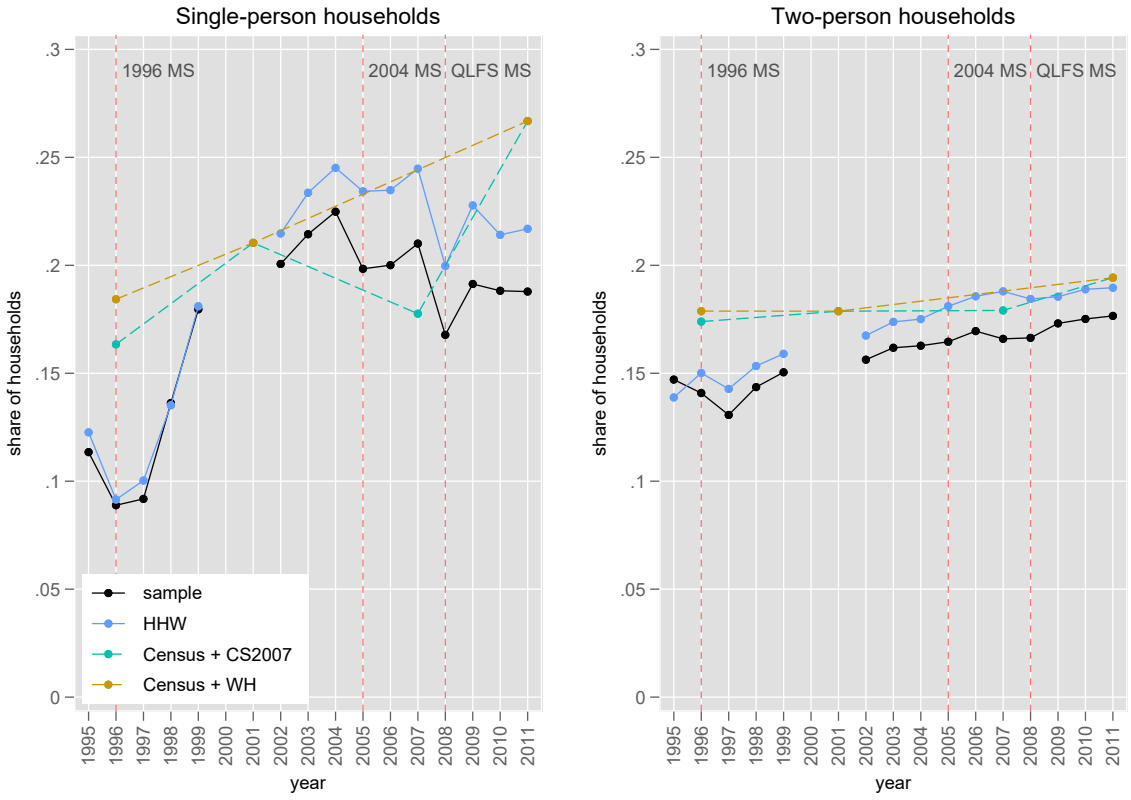
4.3.1 Benchmark quality and scope

There are two sets of possible benchmarks for the survey weights plotted by the 'Census+CS2007' and 'Census+WH' lines in Figure 4.1. The former presents the benchmarks that appear in the StatsSA Headship Model: the three ten percent household samples from the 1996, 2001, and 2011 censuses, and the 2007 Community Survey (CS2007). As we have already mentioned, the CS2007 is not a proper census; but instead, a larger-than-usual household survey making it prone to the same pitfalls that plague the sampling of any household survey. The CS2007 estimate for single-person households stands out for appearing too low. This is particularly noticeable because the estimate for 2007 comes in lower than the estimate for 2001, at 18 and 21%, respectively. Inspection of trends in other key outcomes in Chapter 3, such as household size (Figure 3.3) and household counts (Figure 3.2), also shows that the CS2007 is out of step with the census and having an undue influence on the trends produced with our new cross-entropy weight, often causing the trend to dip down between 2001 and 2011 to accommodate its lower benchmark.

As a result, we consider the three actual census estimates to be the more trustworthy indicators of the true trend. However, we also need to deal with the fact that the 1996 census household sample excluded worker hostels, while the 2001 and 2011 samples included them. This is an important omission for the study of small households as we have already explained, particularly for this early part of the post-apartheid period. Both the OHS and the GHS sampled people living in worker hostels making it

[3] Machemedze et al. (2007) recalibrate the weights for the OHS period by constraining the weights on the total number of one-, two-, and three-person households; their new weights adjusted upwards the trends for one- and two-person households, whilst leaving the trends for three-person households relatively unaffected.

Figure 4.1: Measuring single- and two-person households in the census and OHS-GHS series, 1995-2011



Notes: own calculations using the OHS-GHS series, the census ten percent household samples for 1996, 2001, and 2011, and the 2007 Community Survey. sample = unweighted estimates in the OHS-GHS series; HHW = household weight in the OHS-GHS series; CS2007 = Community Survey 2007; WH = census 1996 estimate inflated to account for worker hostels. Vertical lines correspond to change in Master Sample (MS).

appropriate that benchmarks used to calibrate their weights, represent this sub-population. As such, the Census+WH trend excludes CS2007, but inflates the 1996 estimate to include worker hostels following a similar adjustment by Machemedze et al. (2007). We explain how we reach this estimate in the next section, the purpose for now is to show the difference between the benchmarks in the StatsSA Headship Model and the benchmarks we eventually settle on for our best set of weights. The effect is to increase the share of households that are single-person in 1996 from 16 to 18%.

Although the censuses are the best source available for us to use for household benchmarks, they are not without their own limitations. Three sources of incoherence within the censuses themselves admit inconsistency into our constraints in a way we cannot avoid. Briefly, the first of these has already been discussed in the previous chapter. More than one population count constraint enters the calibration procedure at a time, one coming from the MYPE and one from the census ten percent samples. Since these counts differ slightly, this admits the first form of incoherence. Secondly, the calibration of our weights and StatsSA's calibration of the GHS household weights depend on the StatsSA Headship Model, consisting of 32 age-sex-race household headship rates. However, similar to the dual-weight system, the

weights for people in the census person files almost always differ to the weight for their households in the household file, meaning one can extract various estimate of the same household headship rate depending on which weights one applies.[4] Thirdly, the scope of the person and household files for the censuses differs slightly. Although the person file for all censuses is a ten percent sample of all people in the country on census night (excluding foreign diplomats and their families), all three census releases have followed different rules for who should be in the household file. These different rules, plus what appear to be data cleaning issues in some years, make it unclear to us that the scopes of these two files could reliably be brought into alignment consistently across all censuses. Admitting incoherence into our constraints from these three sources is unavoidable; or, at least the only option is to make a different choice and admit incoherence in a different way. Nonetheless, it is important we are aware of these issues because they ultimately undermine the ability of our weights to meet census estimates precisely for statistics which are calculated even sightly differently to the exact way the statistic enters the calibration as a constraint.

4.3.2 Undersampling of small households

Kerr and Wittenberg (2015) describe how small households were undersampled in the 1994-1998 OHSs owing to fieldworker practise at the time. Fieldworkers were instructed to only sample one household per dwelling unit with probability of selection proportional to size, meaning that smaller households (e.g. domestic workers living on an employer's property or backyard shack dwellers) were systematically missed. The result of this practise can be observed in the black unweighted sample in Figure 4.1. Both one- and two-person households occurred at a much lower rate in the OHS period than they ideally should have, according to either census benchmark. Single-person households occurred at a rate of 16% in the unadjusted 1996 census and 18% when inflating for worker hostels; but, only 9% in the OHS 1996 household sample meaning they were occurring at half the rate they should have by the latter benchmark. Two-person households make up 14% of the 1996 OHS sample, but 17 and 18% of the unadjusted and hostel-inflated census, respectively. From OHS 1999, fieldworker practise was improved by instructing fieldworkers to sample every household on a dwelling unit, amongst other changes (Kerr and Wittenberg, 2015). In accordance with this change, the incidence of single- and two-person households increases and moves closer to the census trend in 1999.

The GHS period began with satisfactory levels of small-household sampling, particularly 2002-2004, when sampling was still based on the Master Sample drawn up from the 1996 census. However, with the change to the 2004 Master Sample in 2005, the incidence of single-person households, in particular, lost ground and the gap between the sample and the Census+WH line widened substantially over time. This happened again with the onset of the QLFS Master Sample in 2008, so that by 2011, single-person households occurred in the GHS at a rate of about 18%, when they occurred at a rate of 27% in the census. Each new Master Sample in the GHS seems to adjust the series further downwards; although, the trend is increasing within each Master Sample, with the exception of the QLFS Master Sample. Both the 2004 and QLFS Master Samples are based on the 2001 census (as is the sampling for CS2007). The 2008 GHS stands out for having particularly poor sampling of single-person households related to a

[4] The separation of person and household weights in the census could possibly be related to the modelling to account for undercounting that happens after the Post-enumeration Survey (PES). The 2011 PES metadata makes it clear that person and household undercounts are modelled separately (even though the scope for both is the same) (StatsSA, 2012*b*). We also found cases in all three censuses of person weights varying within the same households; i.e. weights were not integrated.

higher-than-usual level of sampling errors in this survey, in general (StatsSA, 2008*a*).

Unfortunately, the OHS and GHS household weights (HHW) do not assist in correcting this series and instead mirror the trend traced by the unweighted sample. Looking at the HHW series in isolation, one would struggle to identify that the incidence of single-person households was consistently increasing as in the Census+WH trend line. The HHW series moves incongruously with the Census+WH trend and instead overestimates the share of single-person households in the early GHS, only to seriously underestimate it later on, and muddy the direction of the trend in the process. Overall, Figure 4.1 makes it clear that small households have been chronically undersampled throughout the 1995-2011 period, even if there was a brief improvement in the middle of the period. The StatsSA weights do not assist in overcoming this problem, which could also be aggravated by data quality issues in the benchmarks, themselves. Issues with both the benchmarks and sampling are problems we hope to compensate for as best we can with a careful approach to recalibrating the weights. However, caution must be applied particularly in the case of mitigating the undersampling because, in this type of case, calibration can cause as many problems as it hopes to solve. With these caveats in mind, we turn to describing the weights we construct.

4.4 The weight series

Table 4.1 describes the series of weights that we compare in this chapter. The PW and HHW series are the weights that StatsSA release in the person and household files, respectively, of the OHS and GHS. As discussed in the previous chapter, the PW series is calibrated on person demographics and the GHS HHW series is calibrated on the demographics of household heads. The constraints of the one series never enter the calibration of the other, in spite of both series weighting the same sample of people and households and resulting in a range of conceptual and measurement problems discussed in detail in Chapter 3. The goal of Chapter 3 then, was to construct what we call the CEW1 series in Table 4.1, which combines the same information that feeds into these separate weights into a single weighting procedure to solve these problems.

Now we add an additional three weights that incrementally bring the constraints in the CEW1 calibration more in line with our new objective of accounting for the undersampling of small households and improving the quality of the weights, in general. First, we remove the CS2007 benchmark, which was entering the calibration through the StatsSA Headship Model and used to construct constraints on the demographics of household heads. As we motivated above, the CS2007 is out of step with census trends and appears to be having an undue influence on outcomes of interest, such as household size and the share of single-person households. We remove CS2007 from the headship model when interpolating household headship demographics for intervening non-census years and this yields the series CEW2.[5]

Next we turn our attention to inflating the 1996 census benchmark for our household headship con-

[5]In addition to this update and relevant for all other updates, is the important cleaning issue of how to handle headless or multi-headed households. Over 99% of observations lived in households with one head. For the remaining cases of about 7000 (out of about 1.9 million) people who lived with either no head or multiple heads, we allocated the oldest person to be the head. When there were two people tied for oldest, we chose the oldest male if that was an option. If there were still either zero or multiple heads, the household was sorted in descending order of age and whichever tied-for-oldest household member came first was allocated to be the head. This small number of headless or multi-headed households had a surprisingly destabilising effect on the calibration before they were cleaned.

Table 4.1: Six series of weights in the OHS-GHS series

Series	Description - Constraints
PW	**Person weights in the OHS-GHS** - Individual demographics and province totals
HHW	**Household weights in the OHS GHS** - Non-integrated household weight in OHS (sometimes the HH Head's weight) - HH Head demographics from StatsSA Headship Model in GHS
CEW1	**Cross-entropy weighting (CEW) using StatsSA constraints (from Chapter 3)** - Individual demographics and province totals from StatsSA Mid-Year Population Estimates/DataFirst - HH Head demographics from StatsSA Headship Model
CEW2	**CEW1 with Community Survey 2007 removed from the headship model** - Individual demographics and province totals from StatsSA Mid-Year Population Estimates/DataFirst - HH Head demographics from StatsSA Headship Model with Community Survey 2007 removed
CEW3	**CEW2 updating the census 1996 headship model for worker hostels** - Individual demographics and province totals from StatsSA Mid-Year Population Estimates/DataFirst - HH Head demographics from StatsSA Headship Model with Community Survey 2007 removed and census 1996 inflated for worker hostels
Best CEW	**CEW3 adding constraints for single- and two-person household totals** - Individual demographics and province totals from StatsSA Mid-Year Population Estimates/DataFirst - HH Head demographics from StatsSA Headship Model with Community Survey 2007 removed and census 1996 inflated for worker hostels - Total number of single- and two-person households by race from the census and accounting for worker hostels in 1996

Notes: Individual demographics sourced from: DataFirst (2018) and StatsSA (2018*a*,*c*). Household head demographics sourced from: the StatsSA Headship Model obtained via private correspondence between Martin Wittenberg and StatsSA. Worker hostel correction and single- and two-person household constraints are own calculations using the census ten percent samples.

straints to account for worker hostels, which were omitted from this sample of private residential households only (StatsSA, 1996). Following Machemedze et al. (2007), we use 2001 to inform us about what the levels of worker hostels might have been in 1996. The household headship constraints are based on 32 age-sex-race categories for household heads. We calculate the share of each category who were living in worker hostels in 2001 and use this proportion to inflate the constraints for 1996 accordingly. The biggest revision was to male Black African heads aged 0-34 years whose headship rate increased by 8.4%. To the extent that the share of people living in worker hostels declined between 1996 and 2001, this will be an underestimate. This update yields the CEW3 series.

The changes to the CEW series so far have been tweaks to benchmarks that fit easily into an otherwise

smoothly-working calibration using 103 constraints.[6] Next, we account for the undersampling of small households by adding eight new constraints to bring the total to 111. These new constraints are the total number of single- and two-person households by race of the household head in each census year; that is, four race of the household head categories by two household size categories. Benchmarks for 1996 are adjusted in the same was as before for worker hostels (by using 2001 to compute inflation factors by race of the head and household size) and then interpolated for intervening non-census years. It is important that the constraint depends on race because of the concern that Black African single-person households, in particular, were undersampled more seriously compared to white single-person households in the OHSs (Machemedze et al., 2007). With this update, it was again questionable whether the calibration would compile for the same reasons we were unsure about our first efforts to recalibrate the OHS-GHS series in Chapter 3. The calibration process can fail if too many constraints are being imposed on the data; or, if the requirements of the constraints are too onerous for the calibration to meet at once. Both of these were relevant concerns when adding more constraints to compensate for the serious undersampling of single-person households. Fortunately, the procedure did compile, and this update yields the final series, Best CEW.

4.5 Quality of the calibration procedure

4.5.1 Lambda Output

The new weights introduced in this chapter - CEW2, CEW3, and Best CEW - are calibrated using the same method as CEW1, from Chapter 3. The method is discussed in detail in that chapter and not repeated here. The only point worth reiterating is the role of the λ parameter which is informative about the quality of the calibration and the output of which we discuss now. The weights are calibrated using cross-entropy estimation based on the StatsSA person weights in the OHS and design weights in the GHS. This procedure amounts to solving an unconstrained optimisation problem, the objective function for which is parameterised by λ. Each constraint in the constraint matrix has an associated λ coefficient which is informative about how different the new series of weights is compared to the underlying series on which it was based. Higher values for λ then indicate that a constraint was more informative, meaning the new weights look less like the original weights. This can happen either because the calibration is causing distortion, for example, because the sample is being adjusted to too many constraints. This can also happen if the underlying sample looks very different to the constraints in which case the calibration has to 'distort' it relatively more to bend it into a shape that matches the benchmarks.

Table 4.2 summarises the λ output for the three new weight series. The table presents λ output for all the constraints per year, as well as for categories of constraints. The 'person' category covers the individual demographic and provincial constraints; the 'headship' category covers the household head demographic constraints; and the 'hhsize' category covers the shares of one- and two-person households. We report the average of the absolute value of λ for a given category and the share per category with an absolute value of λ exceeding 5 in parentheses. A λ in excess of 5 is considered high, but only as a rule of thumb (Wittenberg, 2010).

[6] See Chapter 3 for a breakdown.

Table 4.2: Summarised lambda diagnostics for the calibration of CEW2, CEW3, and Best CEW

	Series	**CEW2**			**CEW3**			**Best CEW**			
	Constraint cat.	**person**	**headship**	**overall**	**person**	**headship**	**overall**	**person**	**headship**	**hhsize**	**overall**
1995	$E(\|\lambda\|)$	2.20	2.25	**2.22**	2.27	2.08	**2.21**	3.10	6.00	3.24	**3.94**
	share $\|\lambda\| > 5$	(0.06)	(0.09)	**(0.07)**	(0.06)	(0.09)	**(0.07)**	(0.14)	(0.44)	(0.13)	**(0.23)**
1996	$E(\|\lambda\|)$	1.97	1.66	**1.87**	2.00	1.61	**1.88**	3.26	5.75	3.49	**3.99**
	share $\|\lambda\| > 5$	(0.07)	(0.00)	**(0.05)**	(0.07)	(0.03)	**(0.06)**	(0.14)	(0.44)	(0.25)	**(0.23)**
1997	$E(\|\lambda\|)$	1.95	1.69	**1.87**	1.97	1.64	**1.87**	3.01	4.93	2.87	**3.55**
	share $\|\lambda\| > 5$	(0.01)	(0.06)	**(0.03)**	(0.01)	(0.06)	**(0.03)**	(0.08)	(0.31)	(0.13)	**(0.15)**
1998	$E(\|\lambda\|)$	0.85	1.65	**1.10**	0.83	1.59	**1.06**	1.11	4.88	2.65	**2.31**
	share $\|\lambda\| > 5$	(0.00)	(0.00)	**(0.00)**	(0.00)	(0.00)	**(0.00)**	(0.01)	(0.38)	(0.13)	**(0.13)**
1999	$E(\|\lambda\|)$	1.43	1.50	**1.45**	1.40	1.44	**1.41**	2.59	4.34	2.12	**3.06**
	share $\|\lambda\| > 5$	(0.03)	(0.03)	**(0.03)**	(0.03)	(0.03)	**(0.03)**	(0.04)	(0.31)	(0.13)	**(0.13)**
2002	$E(\|\lambda\|)$	2.03	1.93	**2.00**	2.03	1.93	**2.00**	3.33	7.09	3.48	**4.42**
	share $\|\lambda\| > 5$	(0.01)	(0.09)	**(0.04)**	(0.01)	(0.09)	**(0.04)**	(0.21)	(0.34)	(0.13)	**(0.24)**
2003	$E(\|\lambda\|)$	1.58	2.49	**1.87**	1.58	2.49	**1.87**	3.61	8.54	3.83	**5.05**
	share $\|\lambda\| > 5$	(0.00)	(0.13)	**(0.04)**	(0.00)	(0.13)	**(0.04)**	(0.28)	(0.66)	(0.13)	**(0.38)**
2004	$E(\|\lambda\|)$	1.16	2.37	**1.54**	1.16	2.37	**1.54**	2.58	9.01	4.15	**4.55**
	share $\|\lambda\| > 5$	(0.00)	(0.09)	**(0.03)**	(0.00)	(0.09)	**(0.03)**	(0.04)	(0.78)	(0.13)	**(0.26)**
2005	$E(\|\lambda\|)$	1.44	2.07	**1.63**	1.44	2.07	**1.63**	3.57	9.02	4.16	**5.18**
	share $\|\lambda\| > 5$	(0.00)	(0.09)	**(0.03)**	(0.00)	(0.09)	**(0.03)**	(0.34)	(0.66)	(0.13)	**(0.41)**
2006	$E(\|\lambda\|)$	1.50	2.07	**1.68**	1.50	2.07	**1.68**	3.45	8.67	4.04	**5.00**
	share $\|\lambda\| > 5$	(0.00)	(0.06)	**(0.02)**	(0.00)	(0.06)	**(0.02)**	(0.20)	(0.75)	(0.13)	**(0.35)**
2007	$E(\|\lambda\|)$	1.10	1.91	**1.35**	1.10	1.91	**1.35**	2.76	8.44	3.94	**4.48**
	share $\|\lambda\| > 5$	(0.00)	(0.09)	**(0.03)**	(0.00)	(0.09)	**(0.03)**	(0.03)	(0.72)	(0.13)	**(0.23)**
2008	$E(\|\lambda\|)$	0.90	1.60	**1.12**	0.90	1.60	**1.12**	1.90	7.49	3.85	**3.65**
	share $\|\lambda\| > 5$	(0.00)	(0.00)	**(0.00)**	(0.00)	(0.00)	**(0.00)**	(0.03)	(0.59)	(0.13)	**(0.20)**
2009	$E(\|\lambda\|)$	1.54	2.04	**1.69**	1.54	2.04	**1.69**	2.47	8.48	4.08	**4.32**
	share $\|\lambda\| > 5$	(0.00)	(0.13)	**(0.04)**	(0.00)	(0.13)	**(0.04)**	(0.04)	(0.81)	(0.25)	**(0.28)**
2010	$E(\|\lambda\|)$	0.98	1.73	**1.21**	0.98	1.73	**1.21**	1.80	7.95	3.92	**3.73**
	share $\|\lambda\| > 5$	(0.00)	(0.03)	**(0.01)**	(0.00)	(0.03)	**(0.01)**	(0.03)	(0.75)	(0.25)	**(0.25)**
2011	$E(\|\lambda\|)$	0.84	1.51	**1.05**	0.84	1.51	**1.05**	1.35	7.33	3.72	**3.25**
	share $\|\lambda\| > 5$	(0.00)	(0.00)	**(0.00)**	(0.00)	(0.00)	**(0.00)**	(0.01)	(0.72)	(0.25)	**(0.23)**
Overall	$E(\|\lambda\|)$	1.43	1.90	**1.58**	1.44	1.87	**1.57**	2.66	7.20	3.57	**4.03**
	share $\|\lambda\| > 5$	(0.01)	(0.06)	**(0.03)**	(0.01)	(0.06)	**(0.03)**	(0.11)	(0.58)	(0.16)	**(0.25)**

Notes: own calculations using the OHS-GHS series.

For comparison, the average $|\lambda|$ for the previous chapter's CEW1 was 1.19 and the share of $|\lambda|$ exceeding 5 in the whole calibration was 0.03. This is highly comparable to CEW2 and CEW3 in Table 4.2, for which the mean $|\lambda|$ are 1.58 and 1.57, respectively, and which both have a share of 0.03 exceeding 5. This means that removing CS2007 and updating the census for worker hostels are adjustments that do not seriously unsettle the calibration that was originally set up in Chapter 3. However, the same cannot be said for Best CEW. Introducing the 8 new household size constraints has profound implications for the $|\lambda|$ output. The average $|\lambda|$ more than doubles to 4.03 and a quarter of $|\lambda|$ exceed 5 in the calibration of Best CEW overall. The $|\lambda|$s for all categories of constraint are high, but the headship constraints are particularly so with the mean $|\lambda|$ being 7.2 and as many as 58% of $|\lambda|$ exceeding 5. Throughout the GHS period, the average $|\lambda|$ exceeds 5 for the headship constraints. The highest $|\lambda|$ value in the total calibration was 23.13 and came off the constraint for white male heads in the oldest age category of 65 years and older. Indeed, constraints for older whites in other categories were also very high.

Clearly, adding household size constraints to the calibration has upended what was previously a well-behaving procedure. Does this mean that the Best CEW is a poor-quality calibration? Not necessarily. As explained above, λ can be high because problems with the underlying sample mean the λ are relatively more informative about how to adjust the weights so that estimates of constrained statistics agree with the benchmarks. The undersampling of small households in both the OHS and the GHS puts the Best CEW calibration in exactly this position. For Best CEW, high λ are probably more a sign that there are problems with the underlying sample, which we know to be the case, than with the calibration.

4.5.2 Meeting constraints

It is difficult to draw a conclusion about the quality of the Best CEW calibration form the λ output in Table 4.2 because we cannot tell whether signals about distortion relate to problems with the underlying sample, or poor-quality calibration. Another way to test the quality of the calibration is to assess whether the calibration meets its constraints. Figure 4.2 plots the household head and household size constraints against weighted estimates from the OHS-GHS data to test this. There are 32 age-sex-race household headship constraints for CEW2 and CEW3, and an additional 8 race-household size constraints for Best CEW extracted from the censuses. The constraints are population shares of a given category, e.g. the population share of Black African male household heads aged 0-34 years. To make it easier to view several categories on a single plot, we sum the headship constraints by gender in Panel (a); i.e. sum 16 age-race household headship proportions for men, and the same for women, to effectively plot the male and female headship rate in the total population. We also sum the household size constraints by race in Panel (b) to effectively plot the share of the population living in single- or two-person households. This benchmark only applies to Best CEW, which is why only this weight's benchmark is plotted in Panel (b). Against this, we plot the same statistic estimated with each weight in the OHS-GHS series, as well as, the unweighted version with the 'sample' line to show the raw material the calibration had to deal with. Note we plot CEW1, but don't plot it's headship constraints to avoid overburdening the plot's readability.

To read this figure correctly, one needs to recall that different weight series have different constraints. For example, CEW3 and Best CEW include the update to the 1996 benchmark for worker hostels, but CEW2 does not. As such, CEW2's headship benchmark is plotted in a different line in Panel (a) (dashed

blue line), compared to the headship benchmark for CEW3 and Best CEW (dashed red line). Panel (a) then reports that the constraint that CEW2 needed to meet in 1996, was that 13.82% of the total population were male household heads. Because we incrementally updated the weight series, note that many of the lines in the figure fall right on top of each other. For example, the headship benchmarks for CEW2 versus CEW3 and Best CEW differ in the 1990s, but lie exactly on top of each other between 2001 and 2011 when the constraints for all three weights are identical. Similarly, weight series will also lie on top of each other if they share constraints, which is why CEW2 is hidden underneath CEW3 in the GHS period of both plots in Panel (b). In fact, we want the weight series to lie on top of the benchmark lines, because this means the weighted estimates from the OHS-GHS series are exactly reproducing the constraints.

Generally-speaking, this is precisely the case for the Best CEW trend lines; they lie perfectly on top of its constraints. This exercise is then very reassuring that the calibration producing Best CEW is of good quality, and the high λ coefficients in Table 4.2 are more likely a function of the adjustment to undersampling small households. One exception is male headship in 1998 and 1999 which warrants some discussion. The Best CEW benchmark is that 14.37% of the total population were male heads in 1998, and 14.36% in 1999. Best CEW pins the same figures too high, at 14.44% and 14.55%, respectively. Slight discrepancies by all weights from the benchmarked series are visible for these two years, although not to the same degree as Best CEW. There could be something about the OHS 1998 and 1999 samples that is troubling the calibration across all series.

Another aspect that is notable about male headship is the difference in the benchmarks in the OHS period. An expectation we had was that male headship would increase over time, even in the OHS period, so the steady trend for the CEW3+Best CEW headship benchmark is worth unpacking. The share of male heads in the population was 14.38% in the worker-hostel-inflated 1996 census, and actually declined ever-so-slightly to 14.34% in 2001; but, for all intents and purposes, is effectively a level trend. By contrast, the CEW2 headship benchmark, which includes no correction for worker hostels, increases for male headship as we expected between these years.

The question is, is this a credible adjustment to the benchmark if the change is so sizeable it alters the direction of the benchmarked trend? Close inspection of the headship rates in the 1996 and 2001 censuses reveals that young white heads may be an important factor behind the level trend between these dates for the CEW3+Best CEW headship benchmark. White male heads aged 0-34 years had a total population headship rate of 0.79% (representing 5.7% of male heads) and women in the same category had a headship rate of 0.23% (being 2.76% of female heads) in 1996, uninflated for worker hostels. By 2001, the same figures had almost halved to 0.48% and 0.16%, respectively, whilst headship rates for most other categories increased (or decreased only marginally). If male headship is summed excluding this category of young white heads, then male headship inflated for worker hostels sits at 13.58% in 1996 and 13.85% in 2001 (the uninflated figure for 1996 is 13.02%). In other words, inflating for worker hostels still doesn't push 1996 male headship over the 2001 level. Rather, it is this inflation in combination with the dramatic drop in the share of young white heads in the population that causes this result.

Further, the worker hostel inflation mainly increased the numbers of Black African male-headed households compared to other groups, as we mentioned before. The share of Black African male heads aged 0-34 years in the population was inflated by 8%; and, by 6% for Black African men aged 35-49 years. By contrast, rates for white-headed households were never inflated by more than 1%. The result of the

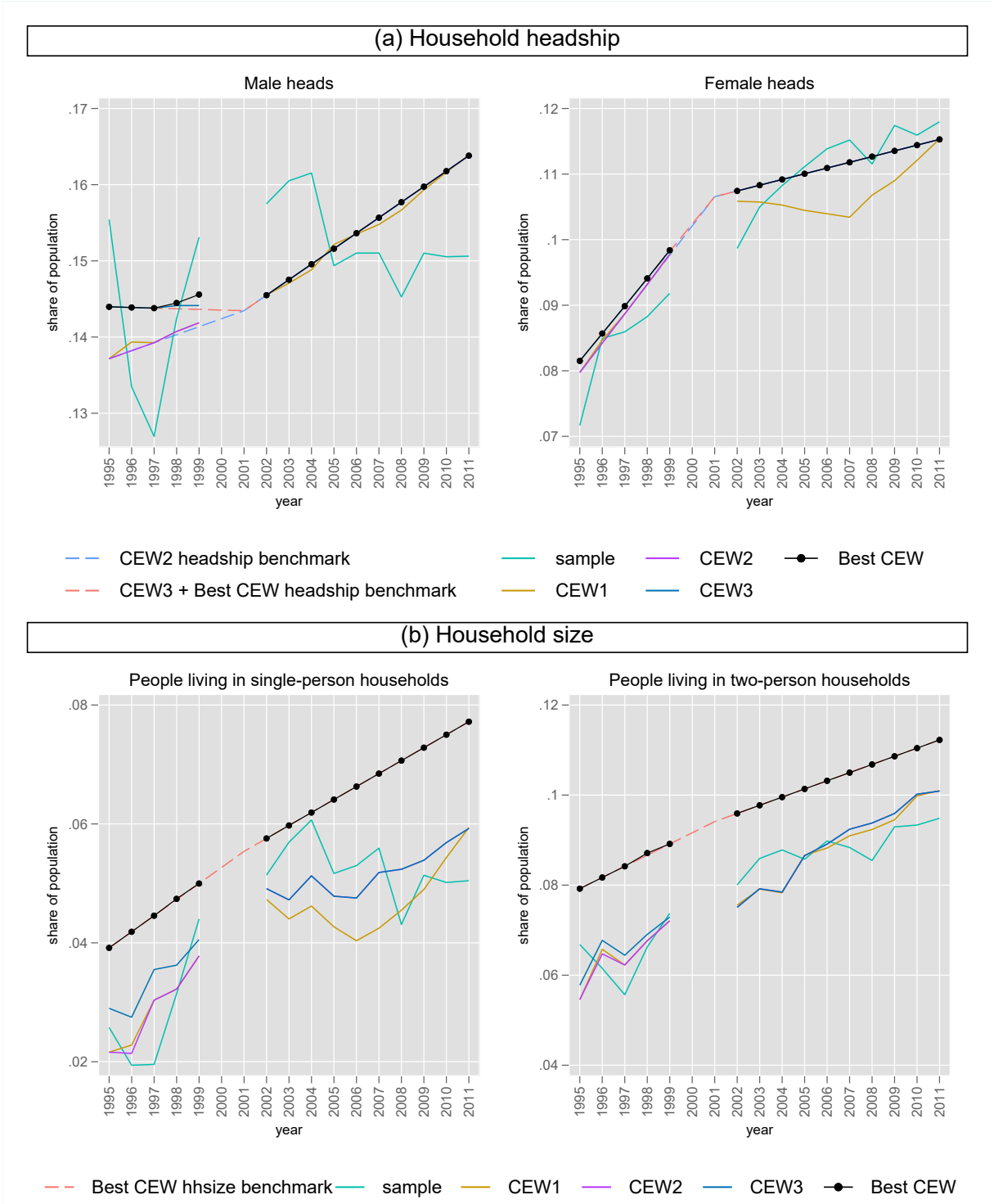
inflation for worker hostels was to increase the rate at which Black African male heads occurred in the 1996 census population from 9.14%, to 9.68%. In the 2001 census, this rate was 10.28%. In other words, the inflation for worker hostels still yields a credible benchmark trend for the share of Black African male heads in the population that is increasing over time.

Understanding these dynamics rationalises the CEW3+Best CEW benchmark. Although the present discussion cannot attribute the drop in the share of white heads in the population to a drop in white household formation per se, compared to other factors, the analysis in coming chapters suggests the former mechanism certainly is at play.[7] It is not surprising that young white people seriously reduced their household formation in this period. The 90's were characterised by much political and economic uncertainty as the country grappled with adjusting to being a new democracy after the first democratic election in 1994. For white people in particular, the end of apartheid meant confronting the end of a legal system that had been set up for their benefit and may have resulted in more conservative household formation behaviour.

Overall, this section has demonstrated that the Best CEW is doing a good job of reproducing its constraints, and certainly as good a job as any of the other weighted series that yielded smaller λ output in Table 4.2. Taking the table and figure in this section together, leads us to conclude that the Best CEW calibration is solid. The Best CEW calibration does, however, have to work harder relative to previous versions of the weight to account for the serious underlying problems with the sampling of small households, which is leading to high λ coefficients in Table 4.2.

[7]This change cannot precisely be attributed to a reduction in the rate at which white people form households in the present analysis because these headship rates could change for numerous reasons owing to the denominator being the total population. The rate at which white people form households could stay the same but their rate of occurrence in the population as a whole could still decline if, for example, the white sub-population is getting smaller relative to other sub-populations. However, analysis in the following chapters, shows white men, especially, curbed their household formation in the 90s.

Figure 4.2: Comparing the CEW weights to household benchmarks



Notes: own calculations using the OHS-GHS series and weight calibration benchmark constraints from the StatsSA Headship Model in Panel (a) and constructed from the ten percent census samples for 1996, 2001 and 2011 in Panel (b). Sample = unweighted estimates in the OHS-GHS series; CEW1-Best CEW = cross-entropy weighted estimates of the OHS-GHS series, with weights and respective benchmarks defined in Table 4.1.

4.6 Results for key demographic outcomes

4.6.1 Population and household counts

Key demographic outcomes for this calibration are person and household counts. Figure 4.3 reports the performance of the weight series for total person and household counts using different weights. For people, the benchmarks comes from the StatSA MYPE in the GHS period and DataFirst (2018)'s population back-projections in the OHS period. For households, we plot both the original set of StatsSA benchmarks (Census+CS2007), as well as, our adjustments to this series (Census+WH). Household counts are also reported in Table 4.3. In Panel (a) for person counts, all weights perform well in matching the StatsSA MYPE except the StatsSA HHW.[8] In Panel (b) for households, it is again evident that the CS2007 is out of step with the other census benchmarks. This means that the greatest improvement in the weight series comes from removing this survey as a benchmark from the headship model and following the Census+WH benchmark. All three CEW series which exclude CS2007 perform well by very closely tracking the census. They also perform almost identically; the series are on top of each other in the plot.

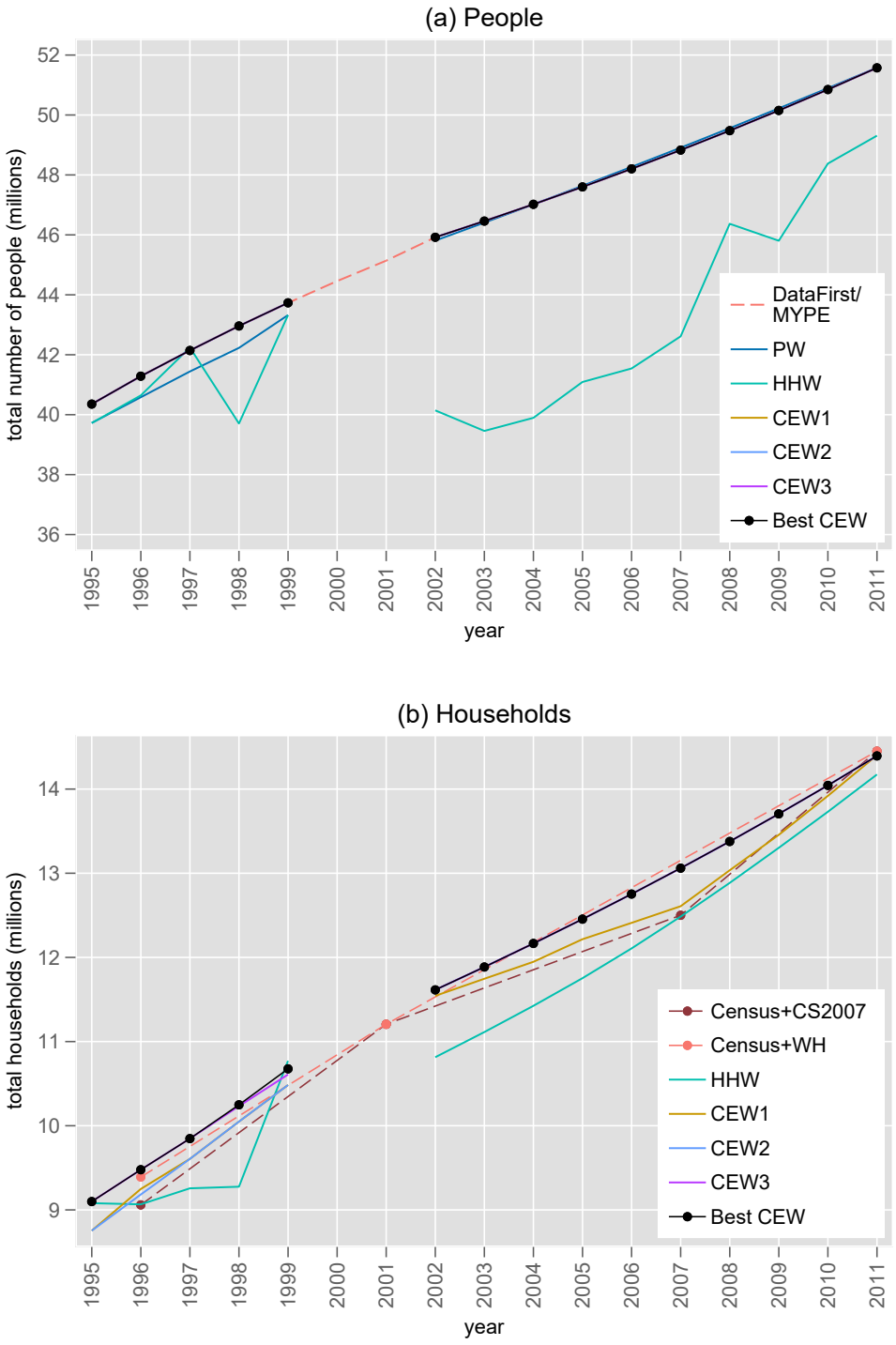
In the OHS period, constraining on worker hostels in 1996 makes a difference to the household counts, visible in the discrepancy between the Census+CS2007 and Census+WH benchmark lines. Both the CEW series that correct for worker hostels in 1996 (CEW3 and Best CEW) report 9.48 million households in 1996 (compared to 9.39 million as the benchmark) in Table 4.3. Slight overestimations of the number of households by all the CEW series is most likely related to unavoidable incoherence within the constraints themselves, discussed in Section 4.3.1, because we have just shown in the previous section that the calibration performs well in reproducing its constraints. Constraining on small households appears to have no additional effect on the estimation of these aggregate outcomes once CS2007 has been excluded and 1996 worker hostels have been incorporated.

4.6.2 Small households

In addition to healthy counts of people and households, the Best CEW also needs to reliably reproduce totals on single- and two-person households. Figures 4.4 plots the counts and shares of single- and two-person households in Panels (a) and (b), respectively. Table 4.4 reports the total counts for single-person households and the two-person version is in Table 4.5. Here Best CEW unquestionably surpasses the other series in performance. The Best CEW series does a relatively good job of reproducing counts in Panel (a), although there is slight overestimation of the order of about 120 000 households at the very beginning and end of the series. Furthermore, Best CEW consistently overestimates the share of single-person households in Panel (b) by about 1-2 percentage points. In 2011, the census reported 26.7% of households were single-person households and Best CEW found 27.7%. These discrepancies are also likely related to the incoherence in the constraints, themselves, discussed in Section 4.3.1. Nevertheless, these slight aberrations can be permitted compared to the performance of the other series. The penultimate best weight, CEW3, consistently underestimates the share of single-person households and the trend line strays further from the census over time in both Panels (a) and (b) of Figure 4.4. By 2011, the StatsSA

[8]The population benchmarks for the OHS period all the CEW series came from DataFirst (2018) which differed slightly from the original StatsSA PW benchmarks. This explains why the PW series is slightly below the benchmark.

Figure 4.3: Total population and household counts in the OHS-GHS series using different weights, 1995-2011



Notes: population benchmarks from DataFirst (2018) (1995-2002) and StatsSA (2018*a*). Census+CS2007 = the StatsSA Headship Model based on the census ten percent samples for 1996, 2001, and 2011, and the Community Survey (CS) in 2007. Census+WH = counts from the StatsSA Headship Model excluding CS2007 and including an own adjustment of the 1996 estimate for worker hostels (WH). Own calculations using the OHS-GHS series weighted: PW = person weight; HHW = household weight; CEW1-Best CEW = cross-entropy weight series described in Table 4.1.

Table 4.3: Total household counts in millions, 1995-2011: comparing different weights in the OHS-GHS series to different census benchmarks

year	PW	HHW	CEW1	CEW2	CEW3	Best CEW	census +CS2007	census +1996 WH
1995	8.44	9.08	8.75	8.75	9.10	9.10		
1996	9.01	9.07	9.25	9.18	9.48	9.48	9.06	9.39
1997	9.23	9.26	9.61	9.61	9.85	9.85		
1998	9.63	9.28	10.05	10.05	10.23	10.25		
1999	10.77	10.77	10.48	10.48	10.61	10.68		
2000								
2001							11.21	11.21
2002	11.48	10.81	11.54	11.61	11.61	11.61		
2003	12.20	11.11	11.75	11.89	11.89	11.89		
2004	12.35	11.43	11.95	12.17	12.17	12.17		
2005	12.63	11.75	12.22	12.45	12.45	12.45		
2006	13.06	12.11	12.41	12.75	12.75	12.75		
2007	13.52	12.49	12.61	13.06	13.06	13.06	12.50	
2008	13.19	12.89	13.04	13.38	13.38	13.38		
2009	13.78	13.30	13.46	13.71	13.71	13.71		
2010	13.49	13.73	13.92	14.04	14.04	14.04		
2011	13.50	14.17	14.40	14.39	14.39	14.39	14.45	14.45

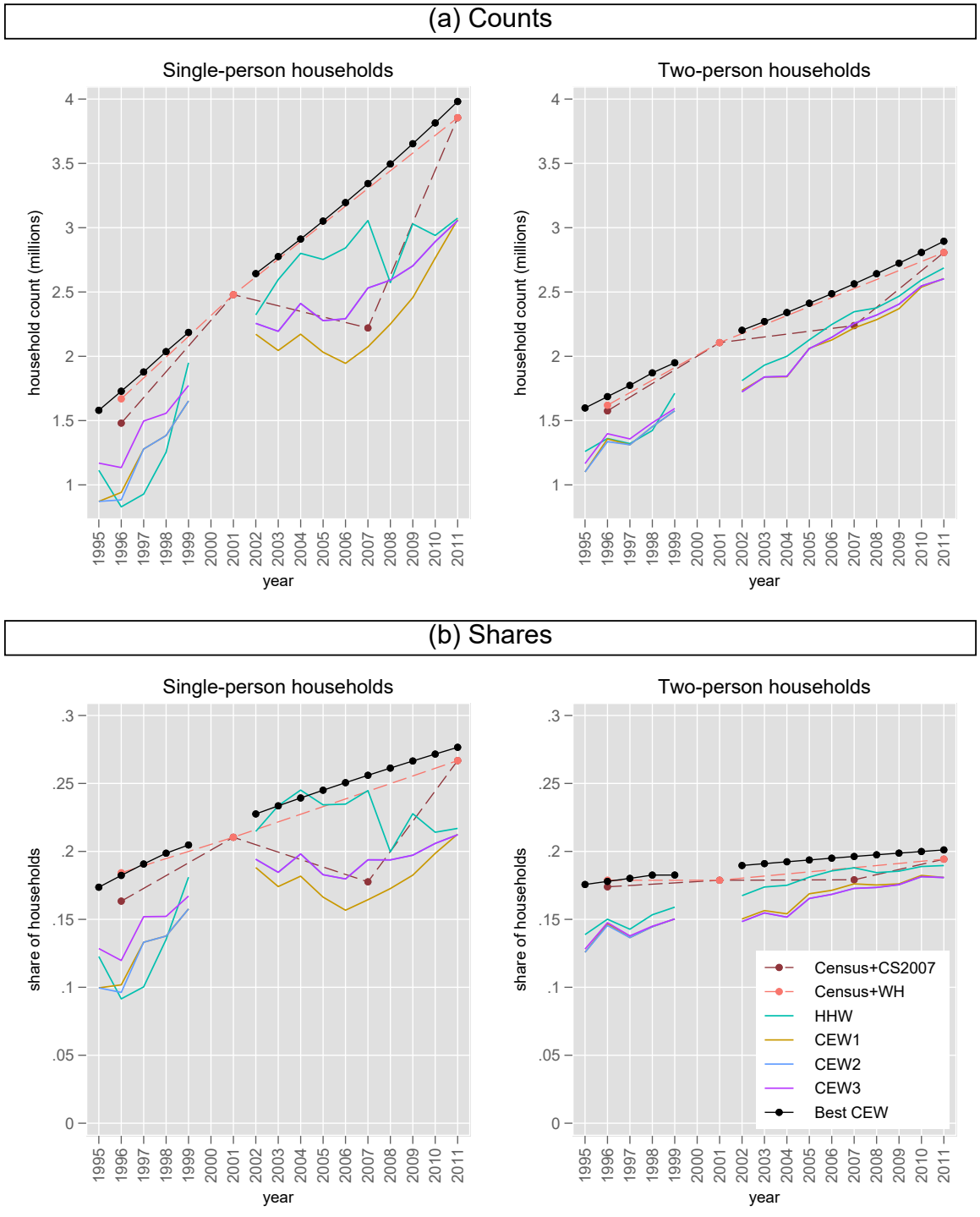
Notes: own calculations. PW = person weight in the OHS-GHS series; HHW = household weight in the OHS-GHS series; CEW1-Best CEW = cross-entropy weights in the OHS-GHS series described in Table 4.1. Census + CS2007 = counts from the StatsSA Headship Model based on the census ten percent samples for 1996, 2001, and 2011, and the Community Survey (CS) in 2007. Census + 1996 WH = counts from the StatsSA Headship Model excluding CS2007 and including an own adjustment of the 1996 estimate for worker hostels (WH).

HHW and CEW3 are underestimating the number of single-person households by the order of about 800 000, or 25%, which is a discrepancy hard to ignore.

4.6.3 Household size

Household size is also an important outcome because part of why the expansion in the household count is of research interest is because of the implications for people's household size and structure. In Figure 4.5, we chart average household size in Panel (a) and the distribution of household size in Panel (b). Panel (a) reiterates what the total person and household counts revealed. Removing CS2007 is the most important adjustment; in fact, its removal makes a material difference to the characterisation of the trend. Outside of this adjustment, all the other weight series perform identically by lying on top of each other and very closely tracking the Census+WH benchmark. The adjustment for worker hostels means that the drop in average household size between 1995 and 2011 is revised down from 0.9 people to 0.7, because average household size in 1995 decreases from 4.5 people according to CEW2, to 4.3 according to Best CEW. Average household size was 3.6 in 2011 according to all CEW series. Once again, constraining on small households appears to make no material difference to the benefits of removing CS2007 and correcting for worker hostels in 1996. This is a good sign, since the small household constraints relate to composition not total person and household counts. What this suggests is that we at least do no harm to our aggregate demographic estimates by choosing the series that also constrains on small households.

Figure 4.4: Single- and two-person households in the OHS-GHS series using different weights, 1995-2011



Notes: own calculations. PW = person weight in the OHS-GHS series; HHW = household weight in the OHS-GHS series; CEW1-Best CEW = cross-entropy weights in the OHS-GHS series described in Table 4.1. Census+CS2007 = census ten percent samples for 1996, 2001, and 2011, and the Community Survey (CS) in 2007. Census+WH = census excluding CS2007 and including an own adjustment of the 1996 estimate for worker hostels (WH).

Table 4.4: Single-person household counts in millions, 1995-2011: comparing different weights in the OHS-GHS series to different census benchmarks

year	HHW	CEW1	CEW2	CEW3	Best CEW	census +CS2007	census +1996 WH
1995	1.11	0.87	0.87	1.17	1.58		
1996	0.83	0.94	0.88	1.13	1.73	1.48	1.67
1997	0.93	1.28	1.28	1.50	1.88		
1998	1.25	1.38	1.38	1.56	2.04		
1999	1.95	1.65	1.65	1.77	2.19		
2000							
2001						2.48	2.48
2002	2.32	2.17	2.26	2.26	2.64		
2003	2.60	2.05	2.19	2.19	2.78		
2004	2.80	2.17	2.41	2.41	2.91		
2005	2.75	2.03	2.28	2.28	3.05		
2006	2.84	1.95	2.29	2.29	3.20		
2007	3.06	2.07	2.53	2.53	3.34	2.22	
2008	2.57	2.25	2.59	2.59	3.50		
2009	3.03	2.46	2.70	2.70	3.65		
2010	2.94	2.76	2.89	2.89	3.81		
2011	3.07	3.06	3.06	3.06	3.98	3.86	3.86

Notes: own calculations. PW = person weight in the OHS-GHS series; HHW = household weight in the OHS-GHS series; CEW1-Best CEW = cross-entropy weights in the OHS-GHS series described in Table 4.1. Census + CS2007 = census ten percent samples for 1996, 2001, and 2011, and the Community Survey (CS) in 2007. Census + 1996 WH = census ten percent samples for 1996, 2001, and 2011, excluding CS2007 and including an own adjustment of the 1996 estimate for worker hostels (WH).

The same is not quite true about the distributional estimates in Panel (b). In this panel, we compare the distribution of household size in 1996 and 2011 in the OHS-GHS series to that in the census household files for the same years. For reasons of space, we only compare the Best CEW to our penultimate favourite, CEW3, because it was the update between these two versions of the weights that made the difference to the household size distribution, which otherwise did not change much. Note also, that due to some inconsistency between the census person and household files, we do not expect a perfect reproduction of the household size distribution in the way that we might of household counts on which the calibration was specifically constrained. Instead, we are aiming for a good approximation of the census distribution, with perhaps the exception of the share of one- and two-person households for the Best CEW, which we have already discussed. A final note, is that we have not adjusted the 1996 census estimates for worker hostels in these plots.

There are a notable discontinuities in the Best CEW distributions around the shares of three-person households, which are clearly too low. The shares of four-person households also look too low compared to the census. Best CEW is only constrained on one- and two-person households because, at the outset and even looking at the CEW3 panel, there was no indication that three- and four-person households were underestimated in the underlying sample. Instead, this outcome is more likely the result of distortion caused by Best CEW compensating for the serious undersampling of single- and two-person households. In increasing the share for one- and two-person households, the calibration compensated by mainly displacing the shares for three- and four-person households more so than other household sizes. Indeed, the shares

Table 4.5: Two-person household counts in millions, 1995-2011: comparing different weights in the OHS-GHS series to different census benchmarks

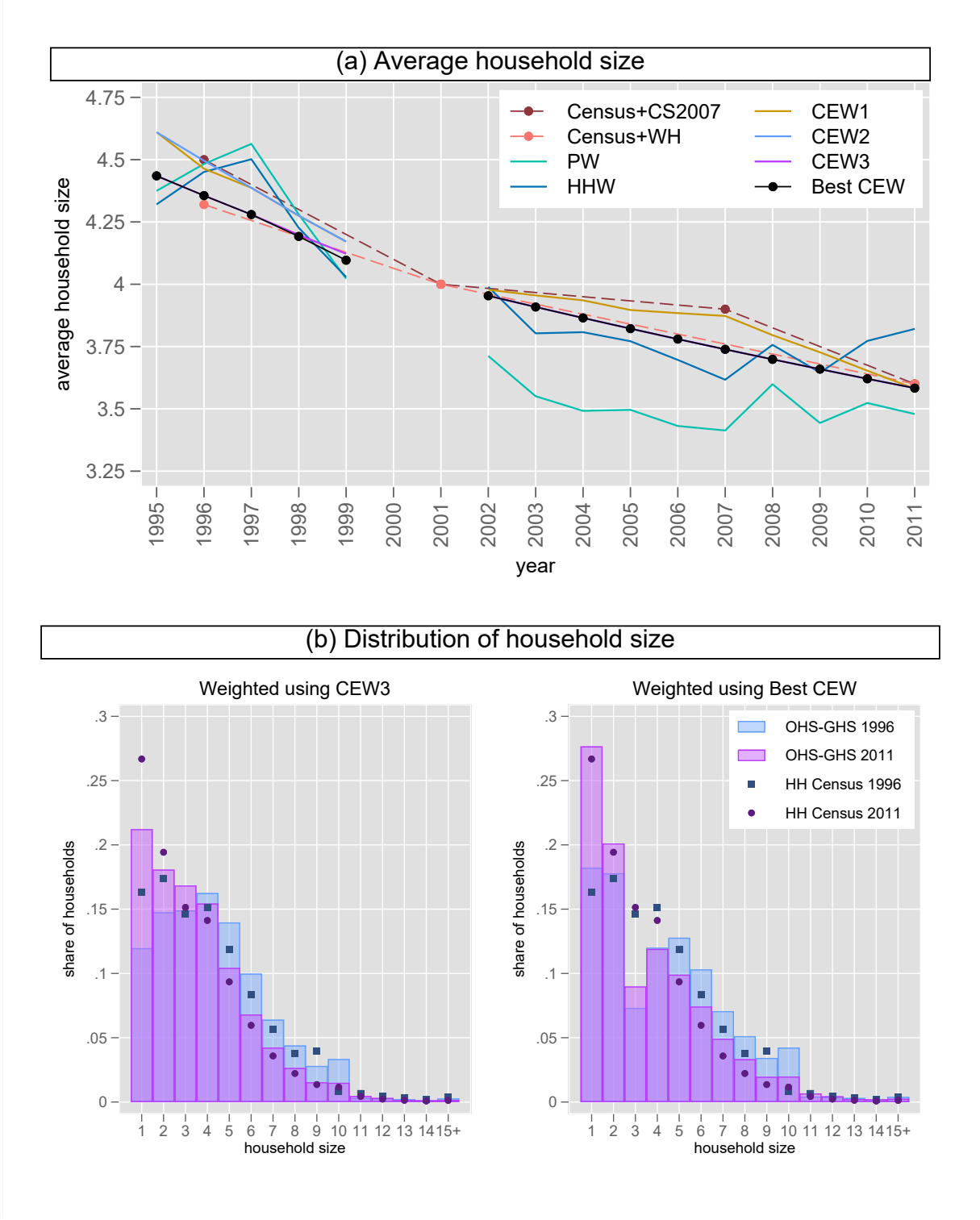
year	HHW	CEW1	CEW2	CEW3	Best CEW	census +CS2007	census +1996 WH
1995	1.26	1.10	1.10	1.17	1.60		
1996	1.36	1.36	1.34	1.40	1.69	1.58	1.62
1997	1.32	1.31	1.31	1.36	1.77		
1998	1.42	1.45	1.45	1.48	1.87		
1999	1.71	1.58	1.58	1.59	1.95		
2000							
2001						2.11	2.11
2002	1.81	1.73	1.72	1.72	2.20		
2003	1.93	1.84	1.84	1.84	2.27		
2004	2.00	1.84	1.84	1.84	2.34		
2005	2.13	2.06	2.06	2.06	2.41		
2006	2.25	2.13	2.15	2.15	2.49		
2007	2.35	2.22	2.26	2.26	2.56	2.24	
2008	2.38	2.28	2.32	2.32	2.64		
2009	2.47	2.37	2.41	2.41	2.72		
2010	2.59	2.54	2.55	2.55	2.81		
2011	2.69	2.60	2.60	2.60	2.89	2.81	2.81

Notes: own calculations. PW = person weight in the OHS-GHS series; HHW = household weight in the OHS-GHS series; CEW1-Best CEW = cross-entropy weights in the OHS-GHS series described in Table 4.1. Census + CS2007 = census ten percent samples for 1996, 2001, and 2011, and the Community Survey (CS) in 2007. Census + 1996 WH = census ten percent samples for 1996, 2001, and 2011, excluding CS2007 and including an own adjustment of the 1996 estimate for worker hostels (WH).

for households sized six and larger appear slightly overestimated by Best CEW. It is not entirely clear why this distortion manifested in this way. This might have been the result of the calibration optimising the household size constraints whilst satisfying all the other constraints on person, province and household head demographics at the same time. This issue could potentially be fixed in future versions of the weights with a more detailed household size constraint.

The CEW3 distributions, by contrast, have no discontinuities. This should not placate us into thinking these distributions are then better representations of the household size distribution, because the shares of one- and two-person households are too low relative to the census. The CEW3 distribution has exactly the opposite problem to the Best CEW distribution. So whilst the discontinuity in the Best CEW distribution is clearly incorrect, it is not clear that we would do better using CEW3. Generally, it appears that, even if the shares in each year differ, the trends all cohere. That is, between 1996 and 2011, households with one to three people grew in proportion, whilst larger households shrank.

Figure 4.5: Household size in the OHS-GHS series using different weights, 1995-2011



Notes: own calculations. PW = person weight in the OHS-GHS series; HHW = household weight in the OHS-GHS series; CEW1-Best CEW = cross-entropy weights in the OHS-GHS series described in Table 4.1. Census+CS2007 = census ten percent samples for 1996, 2001, and 2011, and the Community Survey (CS) in 2007. Census+WH = census excluding CS2007 and including an own adjustment of the 1996 estimate for worker hostels (WH). In Panel (a), census average household size calculated by dividing person-weighted population by the household-weighted household counts. In Panel (b), census distribution of household size calculated using the household files and weights (HH Census).

4.7 Measuring household headship and home-leaving

A final exercise to round out the contribution of Part I is to compare the performance of the difference weight series on key outcomes we use in the remainder of the thesis: household headship and home-leaving. The headship rates in this section differ to the benchmark headship rates previously presented in Section 4.5 by calculating the rate as the number of male heads over the male adult (aged 15+ years[9]) population, for example, instead of the total population. Home-leavers are the share of young adults, aged 15-35, who have left home using a definition adjusted slightly from Simkins (2017*a*). In Chapter 7 on home-leaving we describe this definition of a home-leaver in detail, but do not do so here since it is not germane to the purpose of this chapter. Our goal now is to compare the trends in home-leaving produced by different weight series.

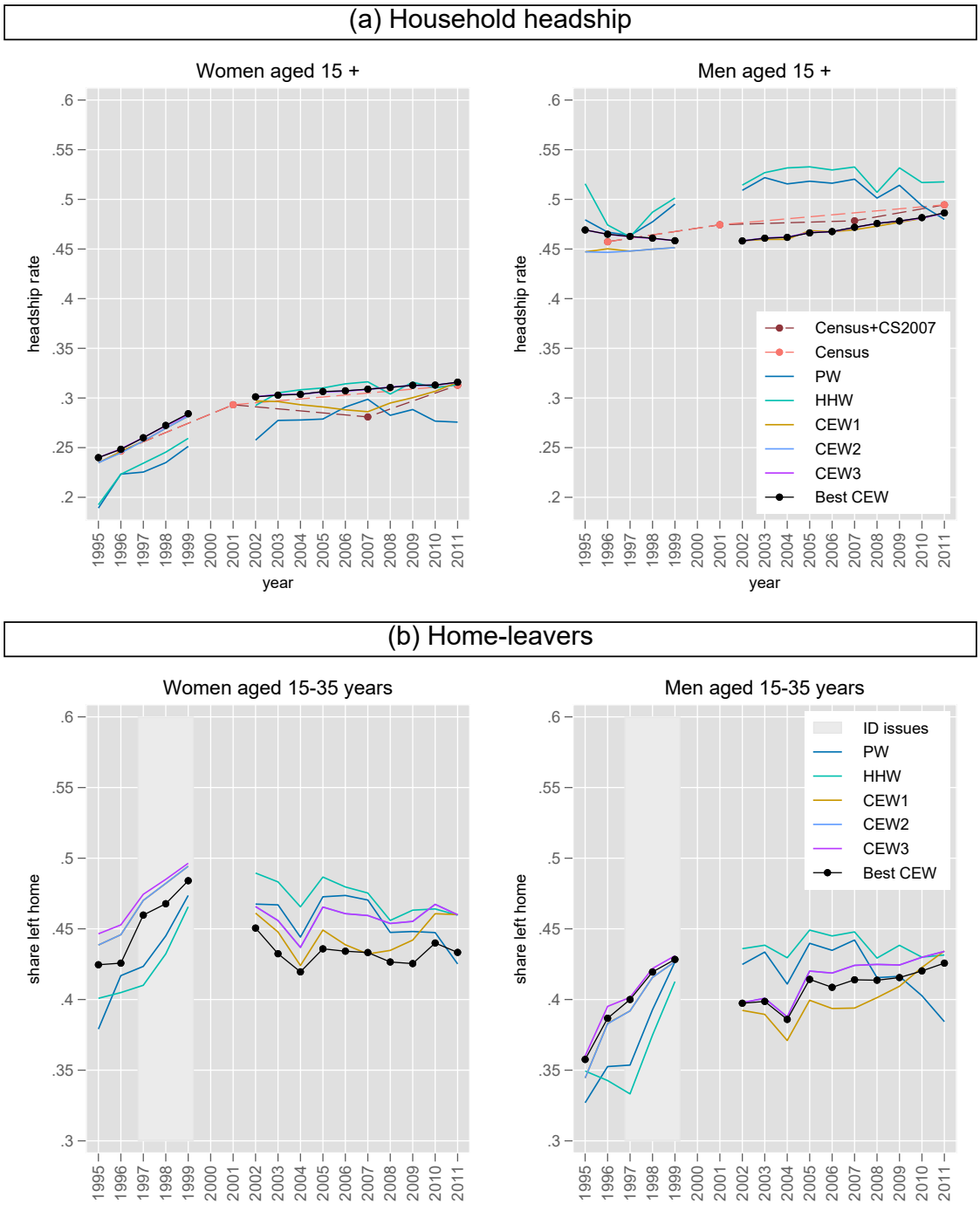
In Figure 4.6, we plot the trend in the headship rate by gender in the OHS-GHS series in Panel (a), and home-leaver status in Panel (b). The only important point to note about the home-leaving variable is that OHSs 1997-1999 lacked variables required to identify home-leavers and these years are greyed out and marked 'ID issues' on the plot, to remind the reader that we can expect home-leaving to be over-estimated in these years for this reason. We don't plot the census in Panel (b) for the same reason of missing variables. The first thing to note about Figure 4.6 is that the Best CEW yields headship trends in Panel (a) that are much smoother than either of the StatsSA weights. The trend for women is clearly improved by removing the CS2007 benchmark which causes a kink in the CEW1 series. Both the census benchmarks in this panel are calculated in the census person files, which should include worker hostels and which is why there is no discrepancy between the Census+CS2007 and Census+WH lines in the 1990s.

For men, the StatsSA weights roughly imply that male headship increases steeply then levels off; whereas, the Best CEW weights suggest male headship slowed slightly in the 90s and then gained momentum in the decade of the 2000s. Although this trend is slightly different in shape to the census, the Best CEW is closer in estimating the level of the census. The difference in trend is possibly related to slight inconsistencies between the person and household files in the ten percent census samples. In Section 4.5, we attributed the downward-sloping trend for Best CEW in the 90s to the inflation for worker hostels somewhat flattening the Black African male household headship trend, in combination with a steep decline in white male headship, who also turn out to be the demographic most likely to head households in this period in the next chapter.

In Panel (b), one might conclude that home-leaving is slowing down if using the PW in the GHS period; whereas, Best CEW suggests female home-leaving is steady in the GHS period and men's is inclining slightly. These are quite different descriptions of home-leaving. All the series appear impacted by the change in Master Sample in 2005, where GHS 2002-4 present different levels and trends to the rest of the GHS series based on the 2001 Master Sample. Constraining on small households makes more of a difference for home-leaving than headship rates. The Best CEW and the penultimate favourite, CEW3, chart trends for home-leaving that differ more than they do for the trends in headship rates. The level of women's home-leaving, in particular, is a few percentage points lower by the Best CEW than CEW3.

[9]We use 15 years as the beginning of adulthood because this is when South Africans are allowed to start working.

Figure 4.6: Household headship and home-leaving by sex in the OHS-GHS using different weights, 1995-2011



Notes: own calculations. PW = person weight in the OHS-GHS series; HHW = household weight in the OHS-GHS series; CEW1-Best CEW = cross-entropy weights in the OHS-GHS series described in Table 4.1. Census+CS2007 = using the person files and weights of the census ten percent samples for 1996, 2001, and 2011, and the Community Survey (CS) in 2007. Census+WH = census excluding CS2007 and including an own adjustment of the 1996 estimate for worker hostels (WH). Headship rate = share of household heads in population aged 15+ years, by sex. Home-leaving = status defined in Chapter 7. ID issues = OHS 1997-1999 lack variables required to define home-leaver status, compromising identification for these years and causing overestimation.

4.8 Conclusion

Serious concerns about the quality of the weights StatsSA releases with the OHS and GHS surveys, such as uncertainty about how to measure the total number of households in a given year, rule them out for use in our analysis of household formation in Parts II- IV of this thesis. The goal of Part I, therefore, has been to calibrate a series of survey weights for the OHS-GHS series that could be considered more reliable than the StatsSA weights. In Chapter 3, we started this process by first paying attention to the most serious problems of the StatsSA dual-weight system and overcoming these by consolidating the multiple StatsSA weighting procedures into one process using cross-entropy estimation. In this chapter, our objective was broadened to improve other aspects of the calibration model that might be undermining the quality of estimates, in general, and for the purposes of studying household formation, specifically. To achieve this, we paid special attention to benchmark choices, but also to the undersampling of small households, which are a particularly important constituency for our study. We made three main adjustments to the developments of the previous chapter: we removed the CS2007 benchmark; inflated the 1996 household count benchmark to account for worker hotels; and, constrained on the number of one- and two-person households.

The outcome is our Best CEW series which is both reliable for cornerstone demographic outcomes, like total person and households counts, but also accurately represents small households. The Best CEW calibration procedure had to work notably harder than previous iterations to compensate for the severe undersampling of small households that occurs throughout the OHS-GHS series. Although the weight matched its benchmarks and outperformed the StatsSA weights on main aggregate outcomes, like household counts, we uncovered some distortion of the household size distribution that serves as an important avenue for future work. For now, the importance of accurate small household counts for our overall research project provides a clear case for adopting the Best CEW to weight our coming analysis in the OHS-GHS series. With the Best CEW, we have a series of integrated weights for the OHS-GHS series which (a) coherently links people to the households in which they live, (b) provides accurate total person and household counts, and (c) provides accurate counts and shares of small households. As such, we adopt the Best CEW in the analysis in the remainder of the thesis.

Part II

Household headship and single-person households

Chapter 5

Trends in household headship and living alone in post-apartheid South Africa, 1995-2011

5.1 Introduction

The goal of this second part of the thesis is to use the data we have just set up to investigate our substantive research interest which is household formation, or proliferation, over the post-apartheid period. Between 1995 and 2011, South Africans have consistently formed households more quickly than the population has grown. According to the OHS-GHS series adjusted with the Best CEW weight, the population grew by 28% between 1995 and 2011, but the total household count changed by 58%; from 9.1 million in 1995 to 14.39 million in 2011, adding 5.29 million new households to the count over 16 years. Of those 5.29 million, 2.4 million (or, 45%) were new single-person households. The number of single-person households mushroomed from 1.58 million in 1995 to 3.98 million in 2011, a change of 150%. At the same time, South Africans were adjusting to the new freedoms afforded by the first democratic election in 1994, and over the span of the period until 2011, key changes occurred in the labour market, marriage market, and social expenditure. Although we know that household size has dropped (Wittenberg et al., 2017), composition has evolved (Posel and Hall, forthcoming), and that women are heading a larger share of households (Posel and Rogan, 2012), we know very little about how trends in household *formation* cohere with these other important changes.

Most studies describing household change in South Africa, focus on compositional change (Steyn, 1995; Amoateng, 1997; Amoateng et al., 2007; Posel and Hall, forthcoming; Seekings, 2008; Russell, 2004). Many of these studies note the drop in household size that has occurred over the post-apartheid period, but few have directly aimed to investigate the process by which this is happening. An exception is a series of studies using data from the Health and Demographic Surveillance System (HDSS) in Agincourt, which

is a rural area in Mpumalanga province where a long-running panel study has been conducted since 1992. Wittenberg et al. (2017) and Wittenberg and Collinson (2020) treat the data as a panel of households and apply a novel decomposition technique to show that household growth in the area is mainly driven by the formation of new households. Wittenberg et al. (2017) reach the same conclusion for the country as a whole, by additionally applying this technique to the NIDS data, which is a nationally-representative panel survey starting in 2008. These studies find that service delivery and household formation are closely intertwined. The provision of free RDP housing is linked to the formation of new households (Wittenberg and Collinson, 2020), and the speed of this process undermined the ability of the State to eliminate the backlog of households that needed to be connected to the electricity grid (Harris et al., 2017).

This chapter extends this literature by providing country-level trends in household formation from the beginning of the post-apartheid period and specifically describing how these patterns cohere with the three most important reasons people form new households in the economic literature: getting older, getting a job, and getting married (Ermisch, 1999; Rosenzweig and Wolpin, 1993, 1994). Our goal is to understand the expansion of the total household count by studying household heads. A focus on heads is motivated by the fact that there is one self-identified head per household in our data, meaning the number of households tallies with the number of heads. Essentially, we conceptualise of the household head as the 'household former', the person in the household most likely to have made the decision to establish their own household at some juncture. We also focus on a sub-set of people heading single-person households as a group of special interest. The study of heads and how they have evolved over time then informs us about patterns and changes in household formation behaviour. Our research question then becomes about describing how South Africa moved from having 9.1 million people forming households in 1995 compared to 14.39 million in 2011; investigating how these groups are different; and how these changes overlap with related trends in the marriage market, labour market, and social spending.

5.1.1 Operationalising household heads

Immediately, we need to deal with the fact that household headship is a contested concept and it is not necessarily obvious that heads can be thought of as 'household formers'. The head of the household is first and foremost a device of survey enumeration. In our data, after the administrative particulars of the household (e.g. PSU number), the very first question asked of the household is to identify the head. Every other household member is then defined in relation to this person. The concept of household head has been challenged by South African and international feminist scholars for being patriarchal; for being an artefact of survey methodology that is abstract from how household members truly understand themselves; and for being amorphous and diversely understood by respondents (Budlender, 2003; Presser, 1998).

Whilst many of the criticisms raised by these scholars ring true, qualitative and quantitative evidence exists for South Africa that the concept of headship is well-understood; that people can easily identify household heads (without asking for clarification of what the concept means) (Rogan, 2016); and that headship is closely associated with being the oldest household member, the breadwinner, and the final authority on household decisions (Posel, 2001). This final quality about decision-making is most pertinent for our purposes of identifying people who at some juncture made the decision to form a new household. Overall, the idea of a single household authority is a familiar one South Africa (Rogan, 2016), where

ideas about hierarchy emerge from patriarchal norms ubiquitous in most cultural groupings (Shefer, 2010; Albertyn, 2009), high levels of religiosity (Chipkin and Leatt, 2011), and these two aspects serving to reinforce messages about male household headship, in particular (Maisiri, 2016).

The connection remains imperfect, though, because headship may not be static[1]; people may inherit headship upon the death of the head[2]; and, people can make the choice to form a household jointly. Ultimately, our choice to focus on heads is guided by our desire to understand the proliferation of households in our census and survey data and focusing on heads allows us to analyse this change directly: the more households we observe, the more heads we observe. As far as we are aware, a systematic investigation of household headship of the type in this chapter does not exist in the South African literature. Most work on headship has focused on verifying the meaning of the concept (Posel, 2001; Rogan, 2016), and another strand of research operationalises sex of the head as a lens on gendered poverty trends (Posel and Rogan, 2012; Rogan, 2013). These studies show that the share of households headed by women increased from 35.2% to 37.5% between 1997 and 2006; that a growing portion of these heads have never married; and demonstrate the higher reliance of female-headed households on grant and female labour market income (Posel and Rogan, 2012; Rogan, 2013). This literature uses headship as a means to understand poverty, but hasn't taken household formation as its main research question.

Household heads in this chapter are also heads of households defined in a strict sense. In our data, respondents are considered household members if they on average slept four nights a week in that household in the four weeks prior to enumeration. This definition is strict given well-established ideas about stretched households and labour migrancy in South Africa (Spiegel et al., 1996) described in Chapter 2. Data constraints mean a limitation of this work is the inability to properly identify labour migrants and gain insight into to what degree changing patterns of migrancy might have contributed to household proliferation over the period. This is particularly an issue for our focus on single-person households, which might be more likely than other household types to be migrant destination households. However, the benefit of the strict definition is that we are able to avoid any double-counting of either people or households; directly tether our results to change observed in census data; as well as describe household change using the definition usually used to compute national welfare statistics.

5.1.2 Living alone: moving beyond census data

A second focus in this chapter is single-person households. Single-person households are of special interest to any study of household proliferation because they contribute so quickly to the household count. As described in the previous chapter, documenting trends in single-person households is of general interest given the global expansion of living alone (Snell, 2017) and of special South African interest given the country's long history of migrant labour (Wilson, 2001). Based on census estimates, living alone has ballooned at an astounding pace in South Africa. In the 1980 census, only 7% of rural and 6% of

[1] Although Klasen and Woolard (2009) find evidence that headship is relatively stable. The African and Indian respondents from KwaZulu-Natal province in the 1993 South African Living Standards Survey were re-interviewed in 1998 for the KwaZulu-Natal Income Dynamics Study. From this two-wave panel they find 96% of household heads or spouses who were alive and resident in 1993 were still head or spouse in 1998. The few 'demotions' from the position of head had an average age of 67.

[2] In South Africa, though, many widows and divorcees are not allowed to remain in their husband's home (Claassens and Ngubane, 2008).

urban households were single-person (Simkins (1986) in Ziehl (2002))[3]; by the 1996 census, this had leapt to 16%, and was up to 27% by the 2011 census (own calculations). In other words, the share of households that were single-person approximately quadrupled in the 31 years between 1980 and 2011, growing robustly by about ten percentage points in two very different periods of South African history as roughly delineated by the 1996 census.

The growth of living alone is clearly an important trend; however, most of the studies in the much smaller literature on single-person households in South Africa rely on census data only (Amoateng et al., 2007; Jhamba and Mmatli, 2015; Mutanda and Odimegwu, 2019; Roux and Geyer, 2017).[4] This is partly because of the known problems with the measurement of small households in the smaller-scale surveys (Kerr and Wittenberg, 2015), which we aimed to mitigate with the reweighting work in Chapter 4. An important contribution flowing out of Part I, then, is being able to use the fifteen data points from the OHS-GHS series to fill in the detail of the years intervening the three census years, allowing us to provide much more information about living alone in South Africa in a way that is also reliable. In general, living alone has been associated with residing in poorer neighbourhoods and in informal or backyard dwellings (Roux and Geyer, 2017). Localised studies of rural areas have not confirmed the expansion in living alone detected in the country-level data (Wittenberg and Collinson, 2007).

This chapter comprises a highly detailed in-depth description of change in household headship and living alone. It begins by describing the change in household counts and running a simple decomposition to show that rising rates of headship and living alone have contributed towards the proliferation of the household count over and above other demographic factors, like population growth, in Section 5.2. Thereafter, in Section 5.3, we describe trends in the rate at which different demographic groups head households or live alone. Sections 5.4 and 5.5 connect these household formation trends to trends in the labour market and marriage market, respectively. We find that faster rates of household headship are more important for explaining household proliferation than changing age structure, in contrast to the West where rapidly aging populations have played a major role in producing more households (Bradbury et al., 2014; Peichl et al., 2012). Regarding who is forming these new households, there are three main narrative points: firstly, women and Black African men are key groups behind household growth. Secondly, the feminisation of the labour market in the 90s coincided with a boom of female headship, although female heads remain more reliant on social grants than male heads. Lastly, we document the rise of the never-married head with a key interaction with the labour market for single-person households and women.

5.2 Has household formation behaviour really increased?

A description of the distribution and change in household headship and single-person households between 1995 and 2011 is in Table 5.1. The table reports the total number of household heads and single-person households in millions and the rate at which heads occur in the adult population (aged 15+ years) in parentheses. Note that single-person households are a sub-set of household heads and are always included in the estimates for household heads. Between 1995 and 2011, the number of household heads

[3]The 1980 census excluded the independent homelands of Transkei, Bophuthatswana, and Venda.

[4]The exception here is Wittenberg and Collinson (2007) who use data from the Agincourt Health and Demographic Surveillance System (HDSS) and Jhamba and Mmatli (2015) who in addition to the 1996 and 2001 censuses, also use the 2007 Community Survey. Chapter 4 suggests single-person households are probably undersampled in this latter survey.

increased from 9.1 million to 14.39, a change of 5.29 million or 58%. The number of people living alone more than doubled from 1.58 million in 1995 to 3.98 million in 2011. The 2.4 million new single-person households make up about 45% of the overall 5.29 million new households, highlighting the importance of this constituency for household growth. The acceleration in the rate at which people are living alone outpaces general household formation. Both the overall headship rate and the rate at which people live in single-person households increased by 5 percentage points over the period. For the former, this represents a 14% increase, whereas in the latter case this represents an 83% increase.

Men head up more households than women, but female headship rates increased by more. Women increased the rate at which they form households by 8 percentage points, compared to just 2 percentage points for men, and as a result contributed as many new households as men to the count by 2011. Men are also more likely than women to head single-person households and also increased their rate of single-living by more than women over the period. This discrepancy already highlights some important gender differentials in household formation over the period. These shifts are tied into household compositional change, which we delve into later on in Chapter 8.

Table 5.1: Counts and shares of household heads and single-person households, 1995 and 2011

millions	**Household Heads**			**Single-person Households**		
(share)	**1995**	**2011**	**Change**	**1995**	**2011**	**Change**
Men	**5.81**	**8.45**	**2.64**	**1.01**	**2.64**	**1.63**
	(0.47)	(0.49)	(0.02)	(0.08)	(0.15)	(0.07)
Women	**3.29**	**5.95**	**2.66**	**0.57**	**1.34**	**0.77**
	(0.24)	(0.32)	(0.08)	(0.04)	(0.07)	(0.03)
Total	**9.1**	**14.39**	**5.29**	**1.58**	**3.98**	**2.4**
	(0.35)	**(0.4)**	**(0.05)**	**(0.06)**	**(0.11)**	**(0.05)**

Notes: own calculations using the OHS-GHS series weighted using the Best CEW. Household headship share = share of household heads amongst the respective population aged 15 years and older. Single-person household share = share of people living in one-person households amongst the respective population aged 15 years and older. The share of single-person households is a subset of the rate of household heads in general.

Clearly, the number of households have increased over time as well as the rate at which people are forming households. However, it is not necessarily the case that more households can be attributed to higher rates of household formation because South Africa's population has also grown and aged over the period. Some new households might be owed simply to a growing population, and some to how the population is distributed over age, since older age groups usually have higher headship rates. Between 1995 and 2011, the population grew from 41.3 million to 51.6 million. The adult population (aged 15 years and older) increased from 26 to 36 million and the average age of this adult population also increased from 35.59 years to 36.36.[5] An important question then is to what extent household proliferation is the outcome of these demographic factors, as opposed to real behavioural change represented by change in headship rates? If we are going to mobilise headship rates as a useful variable with which to understand household formation and change in South Africa, we need to know to what extent headship trends reflect this broader change.

[5]Own calculations using the OHS-GHS series adjusted with the Best CEW weight.

5.2.1 A decomposition method of household change

To separate out these different sources of household growth, we apply a simple decomposition from McCue et al. (2015) who run this decomposition for the United States between 2004 and 2013. McCue et al. (2015)'s method decomposes the change in the total number of households between two time periods into three sources: population growth, changing age structure, and changing headship rates. These authors noted negative headship rates over time, especially for young adults between ages 15 and 34. Rising numbers of households in the United States were instead mainly driven by population growth, but also by the aging of the abnormally large baby-boomer cohort entering smaller retirement-age households. The household count increased but this was mainly attributable to demographic processes.

The decomposition works by holding headship rates from the beginning of the period constant and then projecting forward how many households we would expect in the second period based on how much the population grew and how the age structure changed. The change in headship rates itself is the remainder of the change, interpreted as a 'behavioural change', after these two other purely demographic sources have been accounted for. The equations below describe how each source is calculated:

$$\Delta_{PG} = \frac{H_{t0}}{N_{t0}} * (N_{t1} - N_{t0}) \tag{5.1}$$

$$\Delta_{AS} = \sum_{a} \frac{H_{t0,a}}{N_{t0,a}} * (N_{t1,a} - N_{t0,a}) - \Delta_{PG} \tag{5.2}$$

$$\Delta_{HR} = (H_{t1} - H_{t0}) - \Delta_{PG} - \Delta_{AS} \tag{5.3}$$

where Δ_{PG}, Δ_{AS}, and Δ_{HR} stand for the number of households we would expect based on population growth, age structure change, and headship rate change, respectively, between the two periods, t_0 and t_1. H is the total number of households which corresponds to the total number of heads. N is the total adult population, aged 15 years or older. There are thirteen five-year age categories, a, beginning with those aged 15-19 years and ending with an open category of those aged 80 and older.

Δ_{PG} is calculated by multiplying the headship rate at the beginning period, $\frac{H_{t0}}{N_{t0}}$, by the change in the total adult population, $(N_{t1} - N_{t0})$. This is the number of new households we would expect in t_1 given population growth since t_0, if people continued to form households at the same rate they did in t_0. The first term of Δ_{AS} reproduces this calculation, but for each of the thirteen age bands, and then sums these to reach the combined effect of population growth and change in the age structure. Subtracting the Δ_{PG} yields the number of new households we could expect from change in the age structure net of population growth. The remainder of household change after Δ_{PG} and Δ_{AS} have been accounted for, is assigned to change in the rate at which people form households, Δ_{HR}.

5.2.2 Results of decomposing household change

Results for household heads and single-person households over the total period are summarised in Table 5.2, which reveals a clear role for headship change in contrast to the US case. Of the 5.29 million new households that were added to the household count between 1995 and 2011, 3.5 million (66%) can be attributed to population growth, 360 000 (7%) to changing age structure, and 1.4 million (27%) to change in the headship rate. In other words, household proliferation in South Africa is to some extent the result

of a genuine increase in the rate at which people are forming households over and above demographic factors. This is especially true for single-person households: acceleration in headship rates for both men and women accounted for 73% of new single-person households in total, being 1.76 million more people living alone. This means change in household formation behaviour alone more than doubled the number of single-person households which stood at 1.58 million in 1995. The remainder of the growth in single-person households is owed primarily to population growth. Headship change also appears to have been substantially more important for women than for men in terms of household heads in general. Headship change accounted for almost half of all new households headed by women over the period, or 1.24 million. By contrast, new male-headed households appear mainly the outcome of population growth with headship rate change contributing less than ten percent.

Table 5.2: Decomposing household growth between 1995 and 2011 into three sources, by gender: all households and single-person households

	Men		**Women**		**Total**	
	millions	**share**	**millions**	**share**	**millions**	**share**
Household Heads						
Population Growth	2.17	0.82	1.29	0.49	**3.52**	**0.66**
Age Structure	0.23	0.09	0.13	0.05	**0.36**	**0.07**
Headship Rate	0.24	0.09	1.24	0.47	**1.42**	**0.27**
Total Change	**2.64**	**1.00**	**2.65**	**1.00**	**5.29**	**1.00**
Single-person Households						
Population Growth	0.43	0.26	0.18	0.23	**0.60**	**0.25**
Age Structure	0.03	0.02	0.00	0.00	**0.03**	**0.01**
Headship Rate	1.18	0.72	0.58	0.76	**1.76**	**0.73**
Total Change	**1.63**	**1.00**	**0.76**	**1.00**	**2.40**	**1.00**

Notes: own calculations using the OHS-GHS series weighted using the Best CEW. Household headship share = share of household heads amongst the respective population aged 15 years and older. Single-person household share = share of people living in one-person households amongst the respective population aged 15 years and older. The share of single-person households is a subset of the rate of household heads in general.

Decompositions in general can be very sensitive to choice of beginning and end period, so in order to inspect these results more carefully, we decompose the year-on-year change in the number of households headed by combinations of gender and race. This exercise is only carried out for household heads (including single-person households) because the sample sizes for single-person households alone are smaller, making this more refined exercise less stable. Results are shown in Figure 5.1 for women in Panel A and men in Panel B. The y-axes report the number of new households added in a given year and owed to population growth, change in the age structure and change in the headship rate. These sources are plotted in area plots (which are partly transparent and can overlap with each other) and add up to the total number of new households added in a given year charted by the red dotted line.

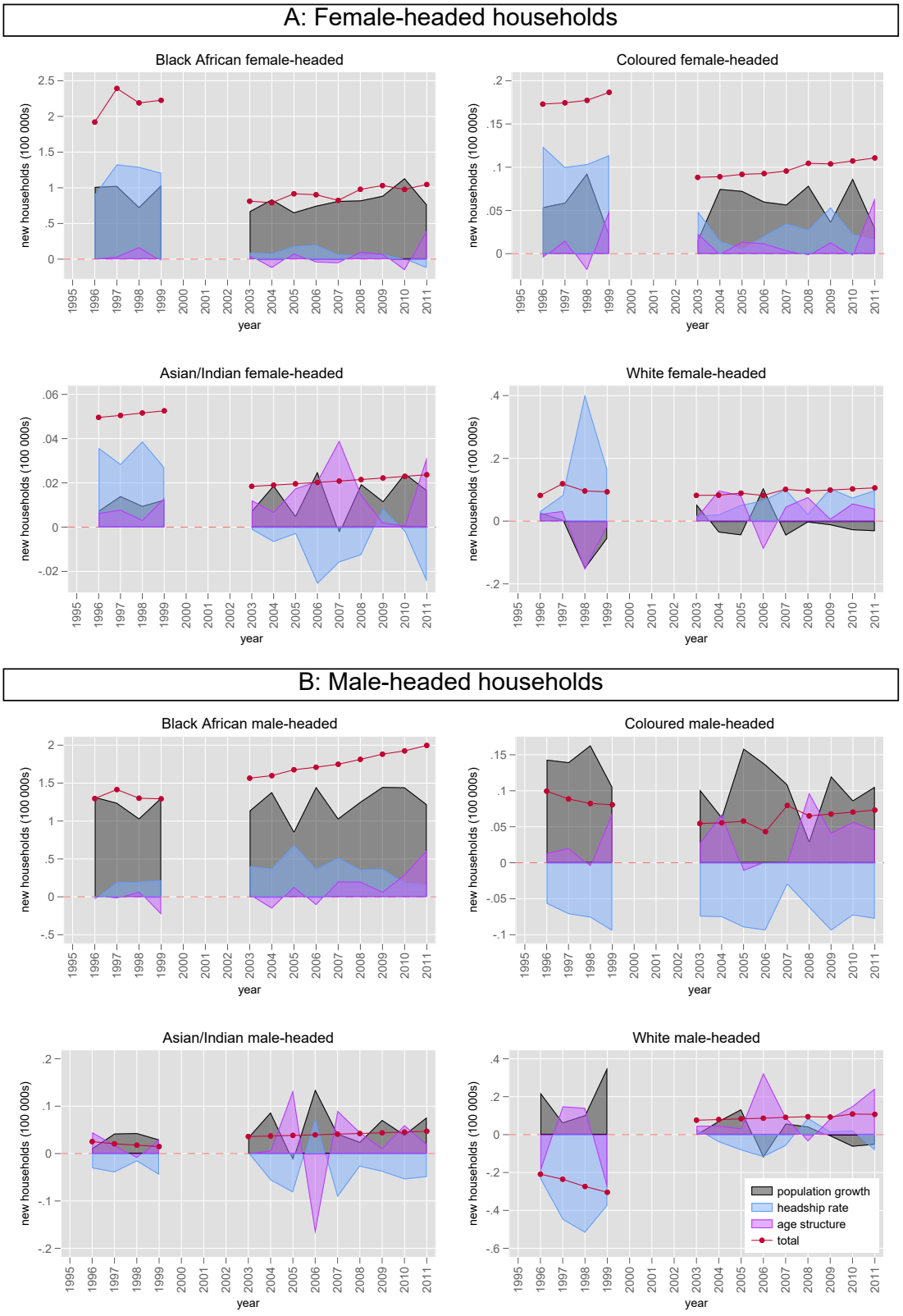
For example, Black African women headed an additional 220 000 new households in 1998 compared to 1997; of this number, about 75 000 were owed to population growth, 125 000 to an increase in the headship rate, and about 20 000 to changing age structure. Note the y-axes are not common across panels because the number of new households contributed by the majority race group, Black African,

will obviously be more than that contributed by the smallest. There is a large discontinuity in the number of new households added by Black African, Coloured, and Asian/Indian women between the OHS and the GHS. This corresponds with the discontinuity in the population and household growth rates at the same point in time in Figure A.1 in Appendix A.1.

Some very interesting gender and race differentials emerge from Figure 5.1. A stand-out result is the large contribution of the female headship rate across race groups in the OHS period. During the OHS period, increasing headship rates were responsible for more than 125 000 of the approximately 225 000 new Black African female-headed households that were added each year in this time. A notable point is the almost complete neutralisation of the headship rate for Black African women in the GHS period, during which new households were almost entirely owed to population growth. The opposite happens for Black African men. Only Black African men increased their headship rates over time while household growth for other groups of men was related to population and age structure effects. The headship rate contributes only marginally to Black African male household formation in the OHS period, but rises in prominence during the GHS, so much so that headship changes for Black African men were more important than for Black African women. In additional calculations using the results in Figure 5.1, headship rate change was responsible for an average of 8% of new households added each year in the GHS period by Black African women; compared to 22% for Black African men.

The takeaway from this section is that a genuine increase in household formation behaviour over and above demographic factors has contributed to the growth of the household count over the post-apartheid period. This is not only different to the developed country case, but also motivates our study of household formation behaviour, and provides some immediate context for gender and race differentials in headship change. Now that we have established a role for behavioural change in the study of household change, we turn to plotting rates of household headship over time.

Figure 5.1: Decomposing year-on-year household growth between 1995 and 2011 into three sources, by gender and race



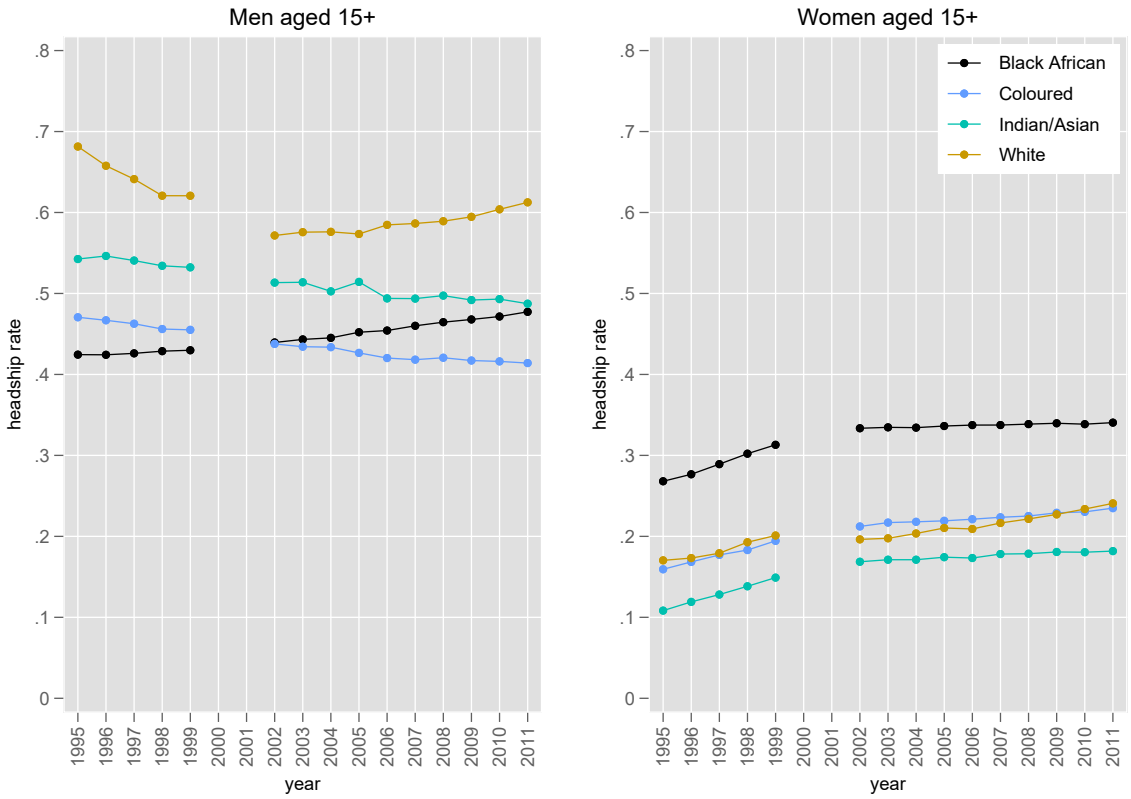
Notes: own calculations using the OHS-GHS series weighted using Best CEW.

5.3 Trends in household headship and living alone, 1995-2011

5.3.1 Household headship

Figure 5.2 plots headship rates by sex and race over time. Men are most likely to be heads, and white men are the most likely. Headship rates for white and Black African men are increasing in the GHS period, but decreasing for Coloured and Indian/Asian men. Note that the increasing headship rate for white men is mainly the consequence of an age effect, as revealed by the decomposition exercise in Figure 5.1. The white male cohort ages up rapidly in this period and older people are more likely to be heads.[6] This means that in effect it is only Black African men who are truly increasing the rate at which they are forming households; a result that also emerged from Figure 5.1.

Figure 5.2: Household headship rates by gender and race in South Africa, 1995-2011



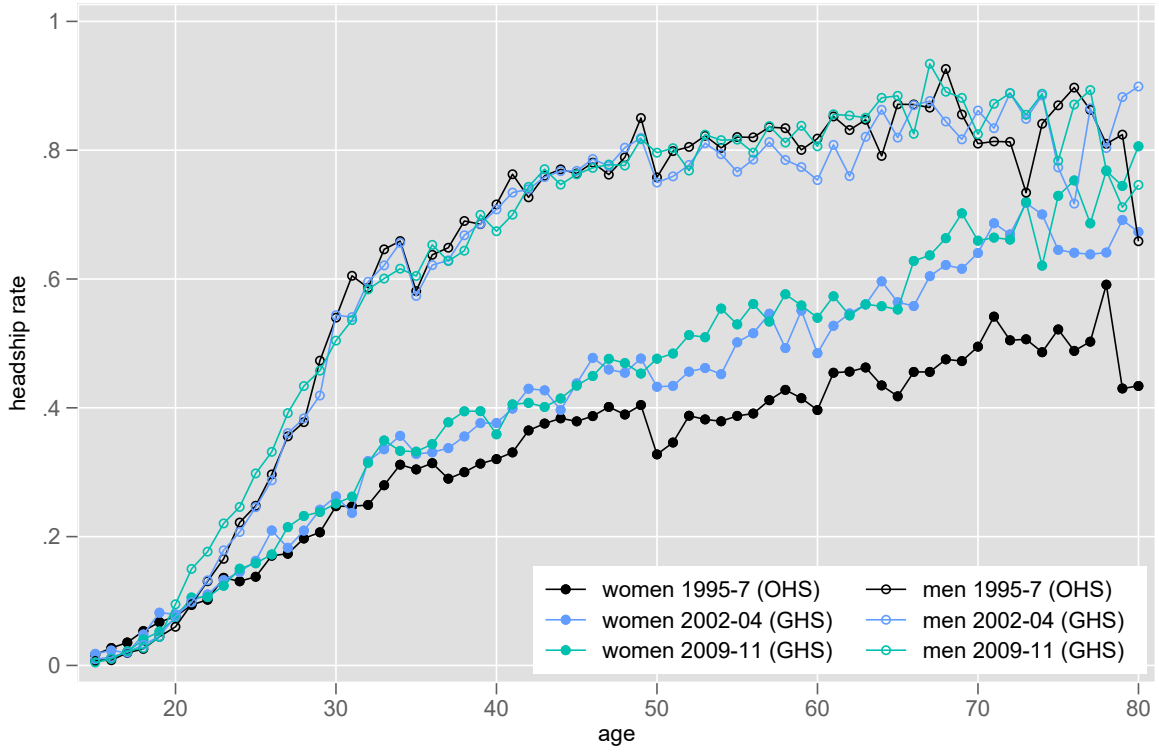
Notes: own calculations the OHS-GHS series weighted using Best CEW. Headship rates calculated as the share of heads over the population aged 15 years and older.

In contrast to men, Black African women are the group who are most likely to head their own households among women. Headship rates surged in the OHS period for all groups of women and this coheres with the important role headship rate change played in the decomposition in Figure 5.1. For Black

[6]The average age of white male adults (aged 15 and older) increases from 40.89 years in 1995 to 44.03 years in 2011 by own calculations using the OHS-GHS series weighted with the Best CEW. For comparison, the same change for Black African men was from 33.62 years to 33.82 years.

African women, headship rates swelled in the late 90s and then plateaued. To a moderate extent, white and Coloured women continue to increase their rate of household formation in the 2000s. The structural break at the turn of the century coincides with an adaption of the age-profile of women's headship rates. Figure 5.3 plots headship rates across age, pooling years in the OHS (1995-7), the beginning of the GHS (2002-4), and the end of the GHS (2009-11) to see when the adjustment in the age-profile took place. The notable change is an expansion in the rate at which prime-aged and older (30 years +) women headed households that specifically happened between the OHS and beginning of the GHS, whereafter the age profile remained stable. The age profile is otherwise remarkably stable for men. Some notable discontinuities in the age profile are at age 35 strongly for men and more weakly for women and again at age 49. It turns out these discontinuities are owed to constraining the weights on household counts (or, in the case of Best CEW, shares of household heads). See Appendix A.2 for detail. Essentially, because the consolidation of person and household constraints into one weighting procedure is so important, we accept these minor discontinuities for the time-being and note them as something to be improved upon in future updates of the Best CEW.

Figure 5.3: Age-specific headship rates by gender, 1995-2011

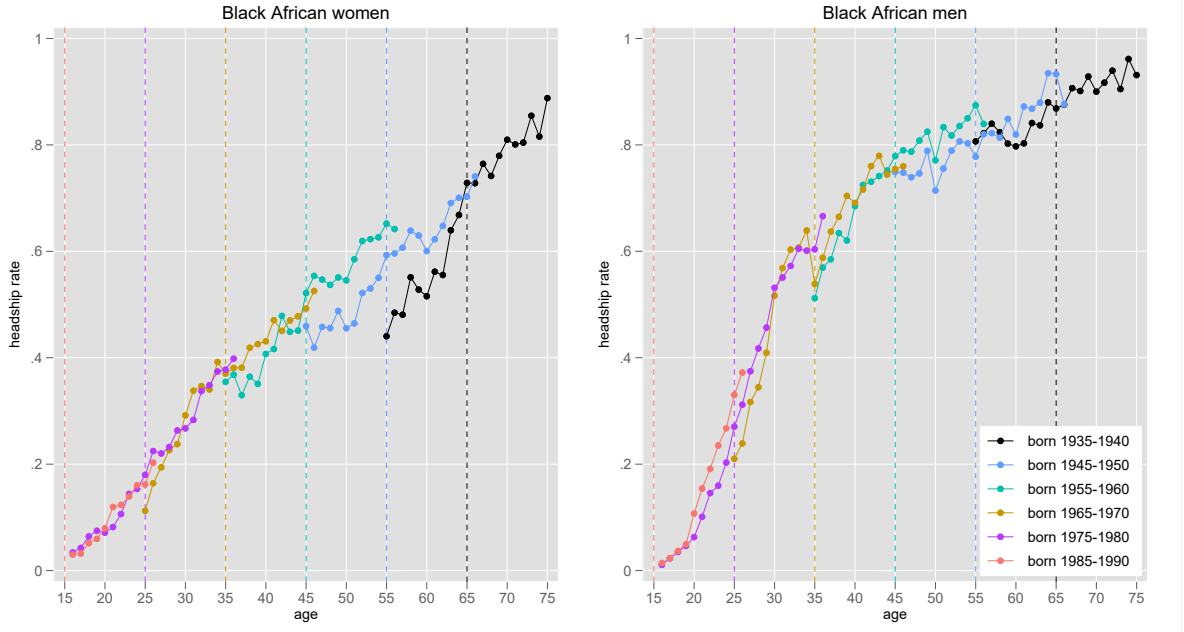


Notes: own calculations the OHS-GHS series weighted using Best CEW. Headship rates calculated as the share of heads over the population aged 15 years and older.

One concern is that the changes in female headship we are observing between the OHS and GHS are driven by this change in instrument. To assess whether this is the case, we plot headship rates for six birth cohorts of Black Africans in Figure 5.4. We focus on Black Africans because their trends are distinct and they represent the majority population group. Vertical lines in the corresponding colour

indicate when the cohort is surveyed in the GHS for the first time. If the change is real, there should be a strong change across generations of women, regardless of survey. Figure 5.4 supports the idea that the change is real; successive generations of Black African women did ramp up the rate at which they were forming households until about the 1965-1970 generation (the gold line). This generation and those coming after it exhibit very similar household formation behaviour suggesting patterns had settled by this point. The expansion of female headship thus happened for generations of women who would have been about 24-29 and older when the first democratic election happened in 1994 (i.e. born 1965-70, the gold line, and earlier). Women who were aged 19 and younger in 1994 adhered to more similar headship rates by age (i.e. born 1975-80, the purple line, and later). By contrast, it is younger generations of Black African men who would have been aged 24-29 and *younger* in 1994 (i.e. born 1960-75, the gold line, and later) who are slightly more likely to form households than their older counterparts, as evidenced by the leftward shift by these younger cohorts between ages 15 and 30. What we cannot tell from this picture currently is whether the expansion of female headship is the winding-down of a pattern that was happening pre-1994, or whether it was a phenomenon of the late 90s.

Figure 5.4: Age-specific headship rates for Black Africans by gender and birth cohort



Notes: own calculations the OHS-GHS series weighted using Best CEW. Headship rates calculated as the share of heads over the population aged 15 years and older. Vertical lines in the corresponding colour indicate when the cohort is first surveyed in the GHS.

A number of important legislative changes discussed in the literature review in Chapter 2 expanded women's freedom in South Africa in the 1990s, giving good reason to think women might have indeed embarked on an expansion in household formation during this time. This leads us to speculate that female headship boomed immediately after apartheid as women in general, but Black women especially, were freed to adjust more fully to marriage, property ownership, and labour market conditions in their household formation choices. This pattern could reflect a once-off adjustment to a new 'steady state' of

household formation constituted by the plateau in the 2000s. Alternatively, the pattern could be the result of optimism that gave way to harsher realities: female household formation was potentially encouraged by optimism in the 1990s owed to extended freedoms and hope surrounding future opportunities in the labour market for more financial independence. But as time wore on, marital prospects worsened and the economy continued to be highly exclusive and unequal, women slowed the momentum of their household formation. We investigate related trends in the labour and marriage markets shortly to better inform these ideas.

5.3.2 Single-person households

The same three plots just discussed for heads (Figures 5.2-5.4) are reproduced for single-person households in this sub-section and show some similarities and differences. Figure 5.5 charts single-person headship rates over time and reveals the opposite ranking by race to the overall rates in Figure 5.2. Black African men are more likely to head single-person households compared to other men; but, white women are more likely to live alone compared to other women. Black African men are responsible for a meteoric rise in male-headed single-person households, whilst single-living amongst Black African women again appeared to surge in the OHS period and level-off somewhat in the GHS period.

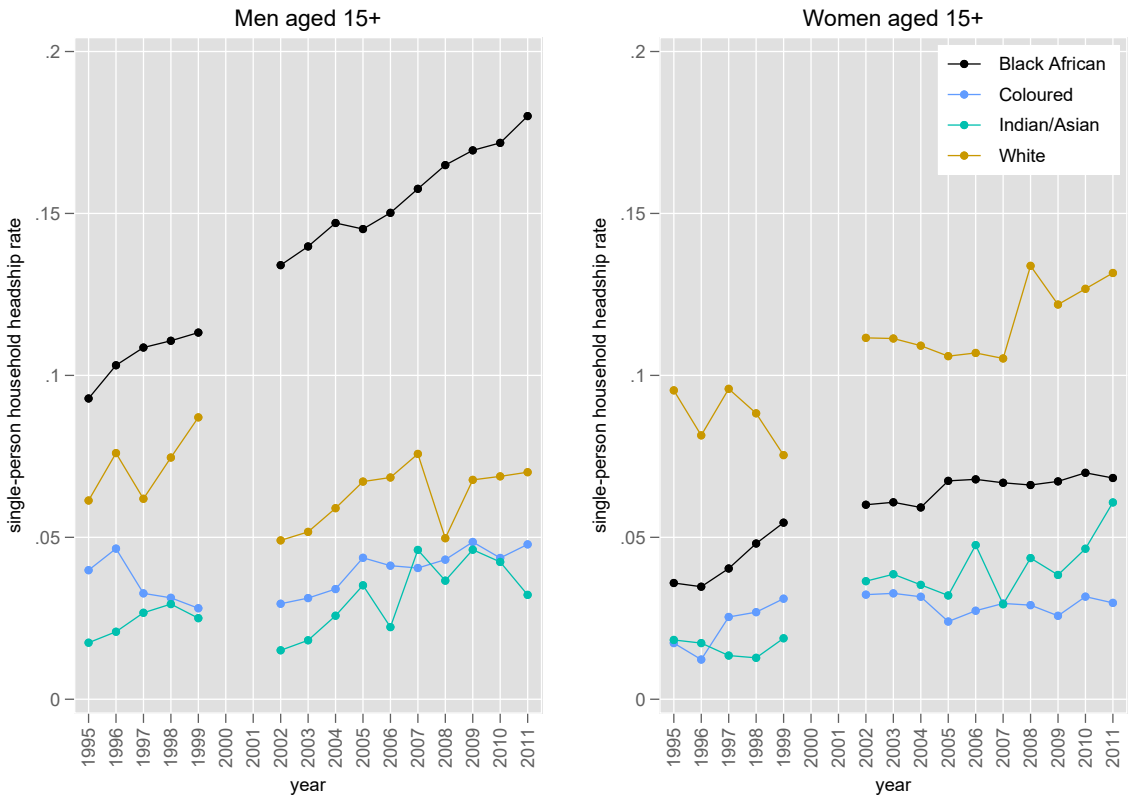
The age-specific headship rates for single-person households are in Figure 5.6. The smaller sample of people living alone makes it necessary to smooth the estimates using local linear regression without which the profiles are very unstable and the plot is not readable. Generally, people become more and more likely to head households as they get older, but the patterns of living alone are more complicated. Single-living rates surge for younger people and then decline over the life course for men, but dip and rise again for older women. These profiles fit the global pattern of gender and age-specific single-living (Esteve et al., 2020; Raymo, 2015).

Increases in the rate of living alone appear to happen at corresponding points in the life course over the time periods. Between the OHS and the beginning of the GHS, the rate shifted upwards most dramatically for young women; whereas in the later period, it was older women who increased their rate of single living most. Similarly for men, first middle-aged men aged 30-50 started living alone more often between the OHS and the beginning of the GHS; and then at the end of the GHS, it was men aged 50 and older who ramped up their rate of living alone. As a result of this process, the risk that senior people will live alone has increased substantially over the period, raising attendant concerns about their well-being. In 2009-11, more than 10% of 70-year-olds were living alone - twice as much as 12-15 years earlier. This rate of change is noteworthy even if rates of seniors living alone remain much lower than the West. More than a third of Europeans aged 65+ lived alone in 2017-2018 (Pleschberger et al., 2019); compared to about 12% of South Africans in 2011 (own calculations).

This distinctive change in age-specific rates of single-living is detectable in pronounced generational shifts in Figure 5.7, again plotted only for Black Africans for the same reasons as before. Men's rate of single-living increases by as much as 8-10 percentage points between generations born in 1945-50 (gold line) and 1955-60 (green line), almost doubling the rate. Discontinuities created by the weights around age 35 are sizeable in this case (see Appendix A.2 for more detail).

In sum, headship rates have increased over the post-apartheid period. Groups driving this change are Black African men; prime and older women in the OHS period; and single-person households. Black

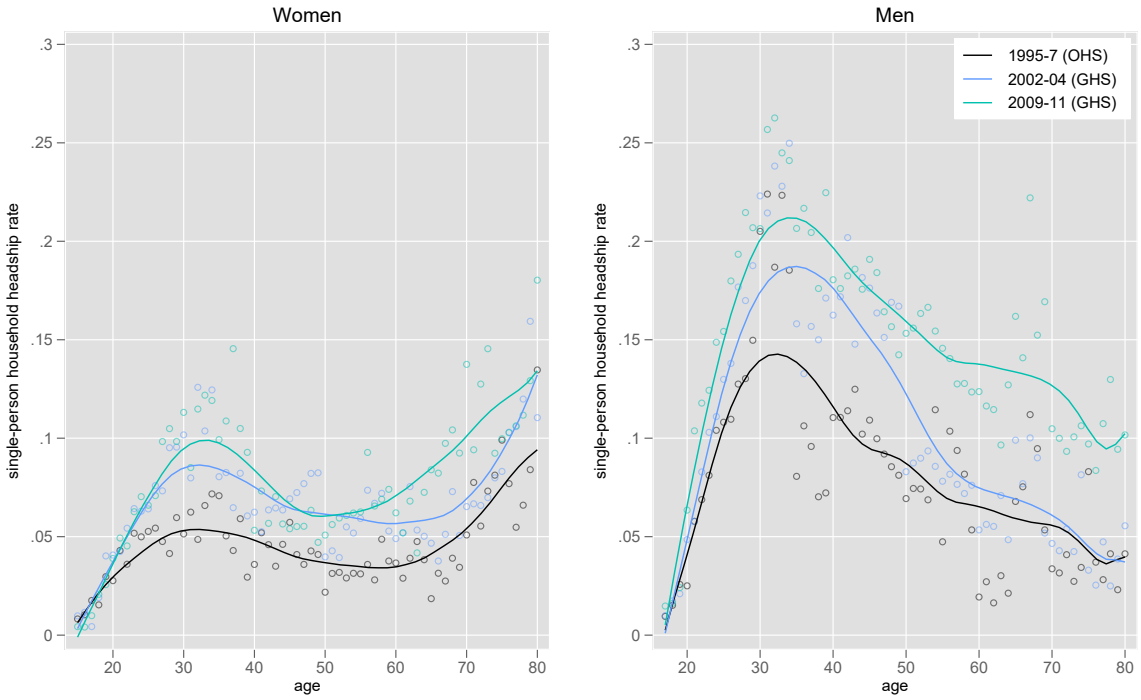
Figure 5.5: Single-person household headship by gender and race in South Africa, 1995-2011



Notes: own calculations the OHS-GHS series weighted using Best CEW. Single-person household headship rate calculated as the share of people heading single-person households, i.e. living alone, over the population aged 15 years and older.

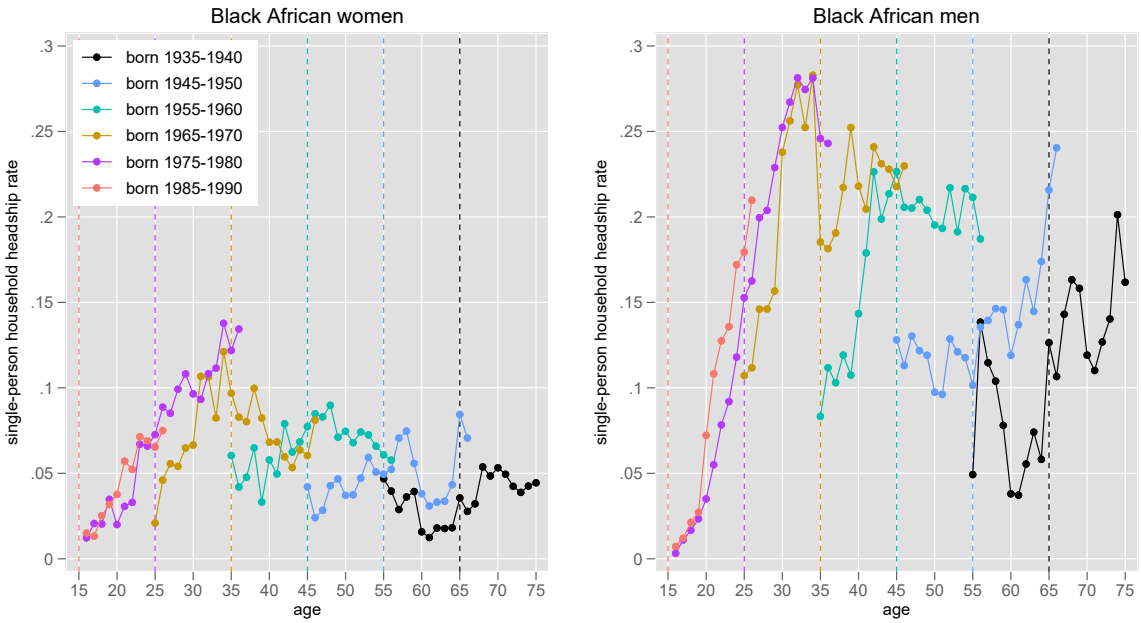
African men followed a slow boil pattern over time; whereas female headship boomed in the 90s and plateaued in the decade of the 2000s. The rate at which single-person households proliferated far outstripped the pace of overall household growth. This was owed largely to consistent increases in single-living by Black African men and younger generations of both men and women, a pattern which seems set to continue beyond 2011.

Figure 5.6: Age-specific rates of living alone by gender, 1995-2011



Notes: own calculations using the OHS-GHS series. Smoothed using an Epanechnikov kernel-weighted local linear regression with a bandwidth of 5 years, weighted using Best CEW. Scatter in the respective color is the share of single-person household headship per age.

Figure 5.7: Age-specific rates of single-person household headship for Black Africans by gender and birth cohort



Notes: own calculations the OHS-GHS series weighted using Best CEW. Headship rates calculated as the share of heads over the population aged 15 years and older. Vertical lines in the corresponding colour indicate when the cohort is first surveyed in the GHS.

5.4 Household formation and gendered access to labour market and grant income

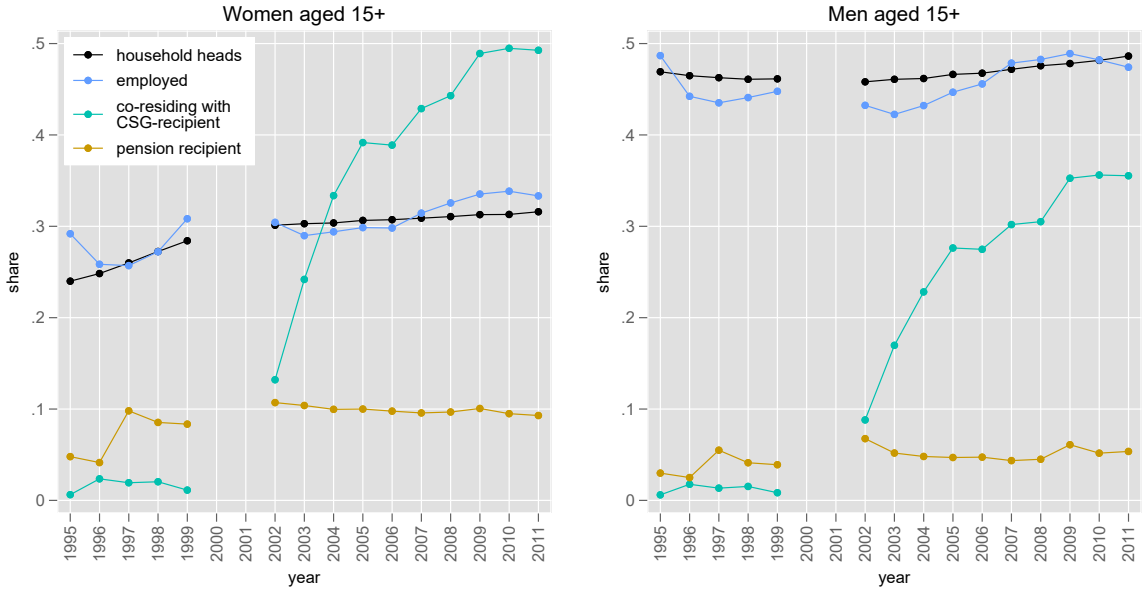
Getting older, getting married, and getting a job are the three most important drivers of household formation that emerge out of the literature on housing tenure and household formation (Ermisch and Di Salvo, 1996; Goldscheider and DaVanzo, 1989). Getting older is associated with strengthening preferences for privacy and autonomy (Fafchamps and Quisumbing, 2007). Getting a job most directly relates to the financial feasibility of forming a new household (Ermisch, 1999) and getting married relates to this pathway insofar as it provides someone with access to their spouses earnings, as well as increasing preferences for privacy for the couple. As such, sketching how household headship evolves in relation to these other trends is of foremost interest. In this section we focus on the labour market, but also on the expansion of the government's social expenditure which is relevant in South Africa's context of high open unemployment.

High levels of unemployment have been a persistent feature of the post-apartheid South African labour market. Male and female strict unemployment rates increased from 17 and 25%, respectively, in 1994 to 23 and 27% in 2011 (Casale et al., forthcoming). This has mainly been the result of more people shifting away from being 'not economically active' (NEA) and finding themselves unable to find work as the economy has struggled to provide jobs. For most of the decade of the 2000s, for example, the share of the working-age population who were employed remained largely constant (Casale et al., forthcoming). Prior to this, the labour market underwent a process of feminisation in the late 1990s as more women entered the labour market after the end of apartheid (Casale and Posel, 2002; Casale et al., forthcoming).

What this means is that the female employment rate followed a boom-plateau pattern very similar to that of the female headship rate. These patterns are visible in Figure 5.8 which charts the employment rate and the household headship rate against time. The extent to which the headship rate tracks the employment rate is quite striking, highlighting the role that access to labour market income has no doubt played in the expansion of female headship. This relates back to the change in the female age-specific headship rate plotted in Figure 5.3. Recall that this figure described how women aged about 30 and older, specifically, expanded their rate of household headship in the late 1990s. This point in the age distribution is significant because when women enter middle-age they enter the years of their life when they are most likely to have a job, partly owing to extreme unemployment rates for female youth. This is described in Figure 5.9, which reports labour market status across the life-cycle with ages 30-55 indicated between reference lines to emphasise women's higher employment likelihood in these years. Also visible in this figure is the expansion of both female employment *and* unemployment across time periods.

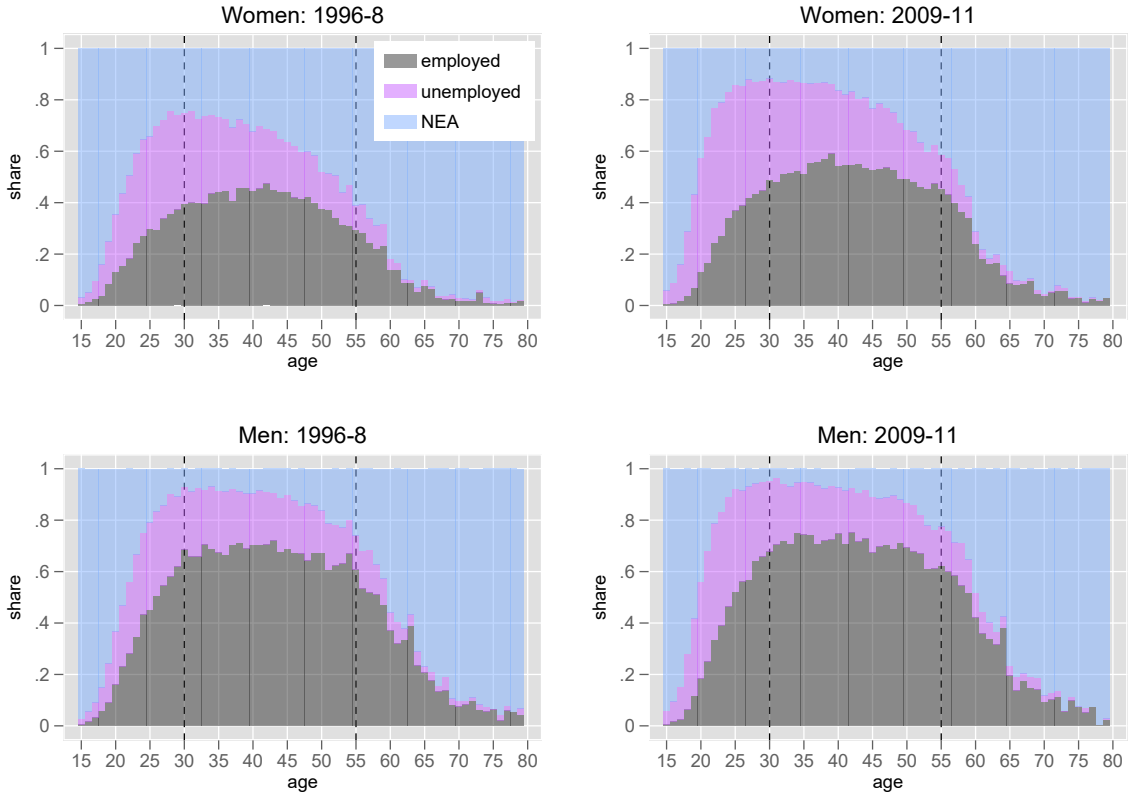
The closeness with which the headship rate follows the employment rate in Figure 5.8 makes it tempting to conclude that the labour market is the definitive determinant of the headship rate. However, Table 5.3 reveals why the story is more complicated. In this table, we report the labour market status of household heads and single-person households over time and show that for women in particular, it is far from clear that access to labour market earnings is synonymous with headship. Less than half of female heads in 2011 were employed at 43.7%, while a similar proportion were NEA. High levels of NEA female heads could be related to a pattern we describe later on where women appear to 'default' into headship upon the onset of widowhood when they are often past retirement age. By contrast most male heads are employed at about two-thirds with comparatively smaller shares being either unemployed or NEA.

Figure 5.8: Trends in household headship, employment, and access to grant income amongst the adult population, by gender, 1995-2011



Notes: own calculations using the OHS-GHS series weighted using Best CEW. CSG = Child Support Grant. The CSG was implemented in 1998, so prior to this year the survey is picking up on child welfare programs that pre-date this social grant called 'maintenance grant/child grant' in the survey questionnaires.

Figure 5.9: Life-cycle labour market status for men and women in 1996-8 and 2009-11



Notes: own calculations using the OHS-GHS series weighted using the Best CEW. Unemployment = expanded definition of unemployment. Reference lines indicate ages 30-55 years, ages when men and women are most likely to be employed.

Table 5.3: Labour market status and access to grant income of household heads, 1995 and 2011

	Men			Women		
	1995	**2011**	**Change**	**1995**	**2011**	**Change**
Household Heads						
Labour Market Status						
employed	66.47	67.12	**0.65**	34.16	43.7	**9.54**
unemployed	13.71	12.93	**-0.78**	19.35	15.91	**-3.44**
NEA	19.83	19.95	**0.12**	46.5	40.39	**-6.11**
Total	**100.00**	**100.00**		**100.00**	**100.00**	
Social Grant Income						
Pension recipient	5.34	9.65	**4.31**	10.27	20.95	**10.69**
Co-resident with CSG-recipient	0.34	21.57	**21.23**	0.66	44.16	**43.50**
Co-resident with other grant recipient	4.80	7.09	**2.29**	6.78	11.71	**4.93**
Single-person Households						
Labour Market Status						
employed	70.52	69.68	**-0.84**	56.99	64.28	**7.29**
unemployed	15.71	15.96	**0.25**	11.12	10.36	**-0.76**
NEA	13.76	14.36	**0.60**	31.89	25.37	**-6.52**
Total	**100.00**	**100.00**		**100.00**	**100.00**	
Social Grant Income						
Pension-recipient	1.43	5.32	**3.89**	3.61	9.20	**5.59**
Other grant recipient	1.44	3.08	**1.64**	2.36	2.05	**-0.31**

Notes: own calculations using the OHS-GHS series weighted using the Best CEW. NEA = not economically active. CSG = Child Support Grant. Other grant = disability, foster care, or veterans grant.

Living alone, though, often means men and women are employed: by 2011, 64% of women and 70% of men living along were employed.

As we mentioned though, many of the NEA heads are likely to be retirees and to better understand the relationship with age we present Figure 5.10. Here, the share of adults per age who are both household heads and of a particular labour market status are plotted separately for men and women and for the beginning and end of our time period. For example, the share of 30-year-old men who are both household heads and employed between 1995-1997 is reported so that if the three plots were stacked on top of each other, they would amount to plotting the total age-specific headship profile for men in that time period. The data are static but labour market status may not be - people can form a household when they are employed and later lose their job but remain head of the household. To some extent, unemployed headship could be reflecting this type of frictional unemployment, although unemployed heads occur at a relatively stable rate of about 10% across age, time periods and genders. Unemployed heads represent a larger share of female than male heads throughout.

There is a distinct life-cycle profile: both men and women start heading their own households in young adulthood when they get a job, until they age up enough to retire when retired people overtake employed people as heads. This happens sooner for women at about age 55 than for men at about age

60. The share of younger men who are employed heads is comparable to the share of older men who are NEA (probably retired) heads; for women though, the share of senior women who are NEA heads far overshadows the share of younger women who are employed heads. Over time periods, not much changes in the labour market profile for men, but for women the expansion of employed and retiree headship is observable across prime and older ages. The results for single-person households are in Appendix Figure A.3 and reflect essentially the same pattern.

Figure 5.10: Headship rates and labour market status over the life-cycle by gender, 1995-7 and 2009-11

Notes: own calculations using the OHS-GHS series weighted using Best CEW. NEA = not economically active. Unemployed = unemployed by the broad definition. Headship rate = share of age group being both heads and of a given labour market status.

Another factor to consider over the post-apartheid period is the expansion of government social expenditure. The provision of a schedule of social grants has been credited as amongst the most effective anti-poverty tools of the post-apartheid government (Leibbrandt et al., 2010). Gender, however, remains a key dimension and higher reliance on social grant income by female- versus male-headed households is one reason why female-headed households stand the higher risk of falling into poverty (Posel and Rogan, 2012). The two most important of these grants are the Old Age Pension which has the highest Rand value and the Child Support Grant (CSG) for its wide reach. South Africa has historically had a non-contributory State-provided Old Age Pension which was initially reserved for whites in the early

20th century before being extended to a limited degree to Black populations (van der Berg et al., 2010). Only in the 1990s was the pension made equally available in terms of both access and amount to the whole population. Women historically qualified for pension-receipt upon turning 60 years of age whilst for most of the period men only became eligible at age 65. However, this changed between 2008 and 2010 when age-eligibility for men was brought down to 60 as well (van der Berg et al., 2010). In 1998, the government implemented the CSG (DSD, 2012). At first, only children under the age of 7 years were eligible and care-givers collect the grant on their behalf. Numerous policy changes have taken place since then to improve access to the grant including increasing the age of eligibility to 18 so that by 2012, about 10.7 million children were recipients (DSD, 2012).

Figure 5.8 also charts the expansion of (adult) access to these two grants. Pension receipt expands in the 1990s and remains stable throughout with about 10% of women and 5% of men collecting pensions. There has been an enormous expansion in the share of adults co-residing with children who are CSG-recipients. In line with what Posel and Rogan (2012) find, women are much more reliant on both pension and CSG income and this is true for household heads too. Table 5.3 also reports the shares of heads and single-person households that have access to pension, CSG and other grant (e.g. disability, foster care, veteran) income. Between 1995 and 2011, the rate at which both male and female heads accessed these grants increased as government's social expenditure increased. In all cases, though, female heads are about twice as likely as male heads to be drawing on grant income. In 2011, 21% of female heads were pension recipients and 44% were co-resident with a CSG-recipient; whereas the same figures for men were 10 and 22%, respectively. The same story plays out for living alone. Just over 9% of women living alone were pension recipients in 2011 compared to 5% of men.

5.5 Household formation and (not) getting married

Declining rates of marriage have been linked to changing household structure in developed countries, where household size has dropped most quickly and single-person households are proliferating most rapidly. In Japan, where single-person households are today the most common household type, changes in marital status were found to almost fully account for their increase (Raymo, 2015). Marriage markets, though are intimately intertwined with labour markets which play a primary role in determining the marriageability of mates (Lichter and Qian, 2019). The process is even further complicated by how gender relations also modulate household headship through marriage. For example, in married couples it is usually men who adopt the headship title (Smit, 2001; Maisiri, 2016). Craigie et al. (2018) bring all these strands together in a study that found that fewer marriageable (employed) men in a local area was significantly associated with higher rates of household headship by never-married women in the United States.

In South Africa, marriage rates have been declining since the 1950s (Garenne, 2016). In our data, the share of women aged 20 and older who are married (civil or traditional marriage) declines from 46.5 to 29.2% between 1995 and 2011; the share who have ever been married (married, divorced, or widowed) declined from 62 to 45.8% over the same period.[7] Notably this is happening in a context where divorce

[7]These estimates are based on own calculations using the OHS-GHS series adjusted using the Best CEW weight. In each case, cohabiting women are not counted as 'married' or 'ever married'. Also note that 'ever married' is only an approximation because the category for divorced stipulates 'divorced or separated' which could be ambiguously interpreted

rates remain low and although rates of non-marital cohabitation have risen, they have not done so to a degree that matches the decline in marital rates (Posel and Rudwick, 2013).

Marital rates also sharply diverge by race, with white women being much more likely to marry than Black African women. In 2010, there was a 40 percentage point difference between the share of Black African (41%) and white (81%) women aged 20 and over who were ever-married (Posel and Rudwick, 2013). The faster decline of marital rates amongst Black African populations has been linked to apartheid policies that separated families and would-be spouses by requiring men to spend long periods away from rural homelands (Smit, 2001; Hunter, 2010). In many cases, working men were housed close to their work-site in single-sex worker hostels which didn't permit women and children to visit (Xulu, 2014). A trend of declining marital rates that began during apartheid has continued into the post-apartheid period and been reinforced by high and sustained levels of male unemployment (Posel and Casale, 2013). High unemployment levels for Black African men in particular have undermined their marriageability and acted as an impediment to paying for bridewealth and the financial commitment of marriage in general (Posel and Rudwick, 2014). A key story emerging from this section is the rise of the never-married head, which seems consistent with Craigie et al. (2018)'s mechanism in South Africa's high unemployment context.

Marital status of household heads and people living alone in South Africa is reported in Table 5.4 for two time periods. There are four marital statuses consistently identifiable in the data: never married, divorced/separated, widowed, and married. 'Married' in this instance includes cohabiting couples and all forms of traditional and civil marriage.[8] We partly follow Rogan (2013) and separate married heads into '*de facto*' and co-resident married household heads. *De facto* heads are married heads living without a co-resident spouse, meaning they could be a migrant worker or spouse of a migrant worker in a stretched household.

Table 5.4 shows that male heads are much more likely to be married with a co-resident spouse than female heads, since men traditionally fill the head position in married couples. *De facto* headship remained consistent at about 8% of male heads in both periods, potentially indicating the continued relevance of labour migration. The predominant change for men was the decline in co-resident married heads in favour of never-married heads. Little change in the incidence of other marital statuses means an almost perfect displacement in terms of shares: the incidence of never-married male heads increased by about 12 percentage points compared to the 13 percentage point reduction in co-resident married heads.

A similar story plays out for women. Never-married and widowed were tied for the most common marital status for female heads in 1995 at about 27% each; but by 2011 never-married female heads were overwhelmingly more likely coming in at 43%. Widowed and divorced or separated heads in 2011 occurred at a similar rate to 1995, meaning never-married heads mainly replaced married (*de facto* and co-resident) female heads. Different to men, though is a much bigger drop in *de facto* female heads. In 1995, *de facto* female heads made up 23% of female heads and co-resident married heads made up 13%. Both categories dropped by about 50% so that the same figures for 2011 were 12% and 6%, respectively. Interestingly, this appears to suggest the opposite conclusion to the one we drew above for men, that there is a drop in the occurrence of stretched households, or at least, those that include conjugal couples. What this might mean instead is that the living arrangement of wives of *de facto* married male heads could be changing.

by people who have never formed a formal union.

[8]The response 'living together' which should identify cohabiting spouses is absent from the 2002-2004 GHS questionnaires.

Table 5.4: Marital status of household heads and single-person households, 1995 and 2011

	Men			Women		
	1995	**2011**	**Change**	**1995**	**2011**	**Change**
Household Heads						
never married	18.66	30.83	12.17	27.24	43.35	16.11
divorced/separated	3.2	2.58	-0.62	8.78	8.1	-0.68
widowed	2.8	4.85	2.05	27.89	30.89	3
married *de facto*	7.83	7.72	-0.11	22.67	11.58	-11.09
married co-resident	67.51	54.03	-13.48	13.41	6.09	-7.32
Total	**100**	**100**		**100**	**100**	
Single-person households						
never married	57.60	71.43	13.83	54.49	66.32	11.83
divorced/separated	11.43	5.41	-6.02	12.42	7.81	-4.61
widowed	5.22	6.53	1.31	21.97	18.19	-3.78
married *de facto*	25.64	16.63	-9.01	10.86	7.66	-3.20
Total	**100.00**	**100.00**		**100.00**	**100.00**	

Notes: own calculations using the OHS-GHS series weighted using the Best CEW. married *de facto* = head is married but no spouse is co-resident. married co-resident = head is married and spouse is co-resident.

A key difference between headship in general and single-person heads is that whilst male heads are most likely to be married, men living alone are far more likely to have never married. After never-married, the next most common marital status is widowhood for women, but *de facto* for men. Even in 2011, 16% of men living on their own were married and possibly living apart from their wives as migrant workers. The main change over the period mirrors the change overall: an extensive increase in the share of never-married heads in conjunction with a decline in married (*de facto)* or ever-married (divorced/separated and widowed for women) categories.

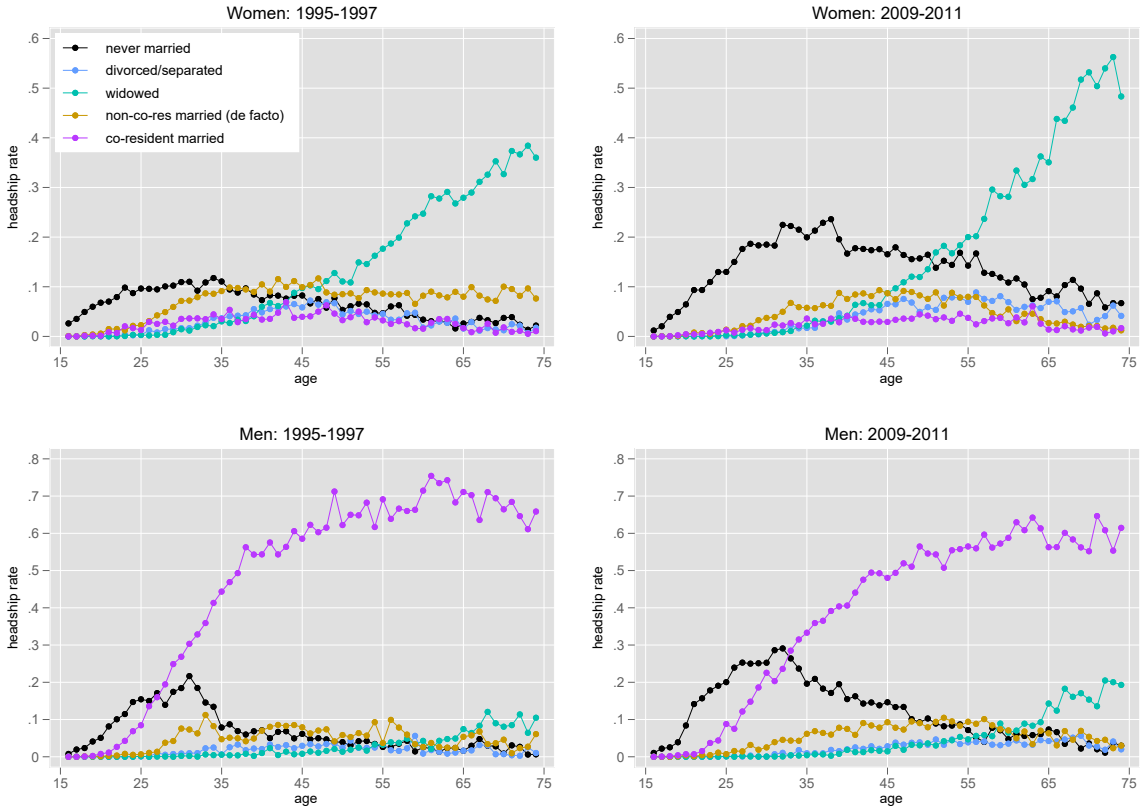
To illustrate how these patterns play out over the life-cycle, Figure 5.11 plots headship rates coincide with different marital statuses across the age distribution at the beginning and end of the period. Similar to Figure 5.10 for employment, in this case we plot the share of 30-year-old men who are both never-married and household heads in 1995-1997, for example. This exercise is only an approximation since marital status is not static and someone could form a household and then marry or get divorced but remain a head.

What is immediately evident from Figure 5.11 is how strongly gender and marital status modulate household formation. Women are typically household heads at a young age before they marry and then 'default' into headship at a later age upon the onset of widowhood. In the middle of their lives, some married women head their own households and it is most common for married women to be *de facto* heads. For men, the typical pattern is to head households when they are young and never-married and gradually this gives way to being married (co-resident) heads as they age up, with some men being widower heads much later in life. We draw these conclusions cautiously given our data is cross-sectional, but this seems like a reasonable reading of the data.

Between time periods, there is a remarkable increase in headship by never-married men and women,

also reflected in Table 5.4. This expansion is most extreme between the ages of 25 and 30, but notably continues into later life as those who had not formed a union in 1995-7 had presumably still not done so by somewhere between ages 37 and 46, 12-16 years later in 2009-2011. The results for single-person households are in Appendix Figure A.4, showing the importance of *de facto* headship for men living alone.

Figure 5.11: Household headship and marital status over the life-cycle by gender, 1995-7 and 2009-11



Notes: own calculations using the OHS-GHS series weighted using Best CEW. non-co-res married (de facto) = household head is married but not co-residing with a spouse. co-resident married = household head is married and spouse is co-residing. Headship rate = share of age group being both heads and of a given marital status.

The changes we observe in Table 5.4 and Figure 5.11 could be the result of two different factors: the changing marital status composition of the adult population and change in the rate at which people with different marital statuses form households. We already know that over time, the share of adults who are never-married has increased as marital rates have declined, but we are also interested in investigating behavioural change. As mentioned above, there is a complicated interrelation between marriage markets, labour markets, gender and household headship. To understand these relationships better, we report conditional chances of headship and single-person headship for combination categories of 'employed-not employed' and 'married-not married'. 'Not employed' includes unemployed and NEA people; 'Not married' includes never-married, divorced/separated, and widowed people. We then report the probability that employed-not married people for example, will be household heads for the population aged 15 years

and older.

Table 5.5 shows that all the action, so to speak, has been in an increase in the chance that not-married people will be heads, regardless of employment status and gender. Being married is the main factor correlated with men becoming heads, but not being married plus being employed is the main factor for women. This showcases the enduring power of gender norms (even if a male spouse is not a breadwinner, he is still the head of the household) and relates back to the earlier discussion about the importance of the feminisation of the labour market for female headship. This pattern held its ranking over time, but the rates at which not-married people (regardless of sex and employment status) formed households increased whilst rates for married people either remained constant or declined slightly. Not married-employed people of either sex are most likely to be living alone by a long margin, suggesting an important interaction between marital and labour market status. The rate at which married people lived on their own remained stable over time, whilst the chance that not-married people would live alone increased, especially for men.

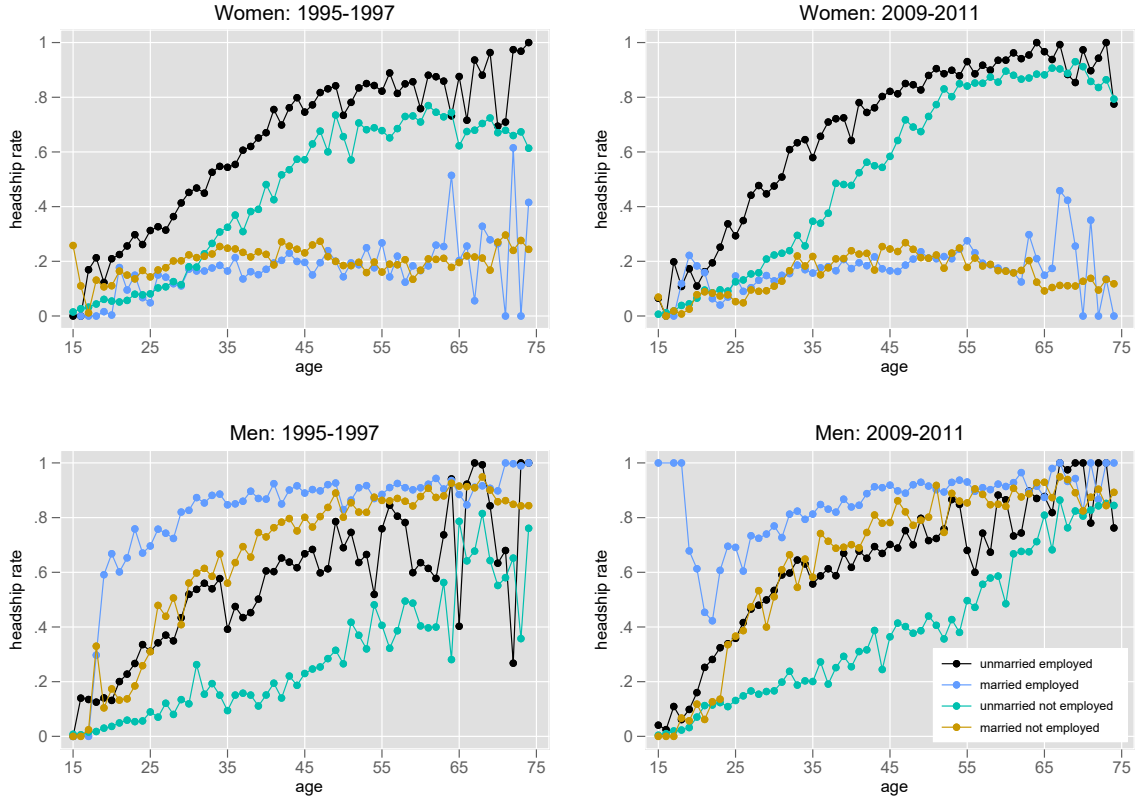
Table 5.5: Probability of heading a household or living alone conditional on labour market and marital status by gender, 1995-7 and 2009-11

	Men			**Women**		
	1995-7	**2009-11**	**Change**	**1995-7**	**2009-11**	**Change**
Household Heads						
Not Employed-Not Married	0.09	0.16	**0.07**	0.22	0.31	**0.09**
Employed-Not Married	0.42	0.50	**0.09**	0.51	0.60	**0.08**
Not Employed-Married	0.80	0.78	**-0.02**	0.21	0.16	**-0.05**
Employed-Married	0.86	0.85	**-0.02**	0.17	0.17	**0.00**
Single-person Households						
Not Employed-Not Married	0.04	0.09	**0.05**	0.03	0.05	**0.02**
Employed-Not Married	0.27	0.37	**0.10**	0.16	0.22	**0.05**
Not Employed-Married	0.02	0.03	**0.01**	0.01	0.01	**0.00**
Employed-Married	0.08	0.09	**0.01**	0.02	0.03	**0.01**

Notes: own calculations using the OHS-GHS series weighted using the Best CEW. Not Employed = unemployed or NEA. Not Married = widowed, divorced/separated, or never married. Married = married or cohabiting.

In Figure 5.12, we chart the rates in Table 5.5 over the age distribution which brings more clarity to the profiles. We report means and impose no functional form on the profile meaning there is some instability at the tails when samples are very small; there are very few 15-year-old males who are married and employed, for example. Even more clearly than in Table 5.5 we can see that not being married is the main correlate for women's headship, regardless of employment status. Prime-aged women who found themselves unmarried but nevertheless employed increased their rate of household formation over time; after age 55, there is a noticeable increase in the rate at which unmarried and not employed (retired) women headed households. Being employed and married is the main correlate for male headship, but failing this combination being either married *or* employed makes men equally likely to head households in 2009-11 until close to retirement age. Over time, unmarried men of either employment status increased their rates of household formation from younger ages than women, from as early as 20. Again, the results for single-person households are in Appendix A.5 and this time confirm the importance of the employed-unmarried combination for both men and women and how this profile has expanded over time.

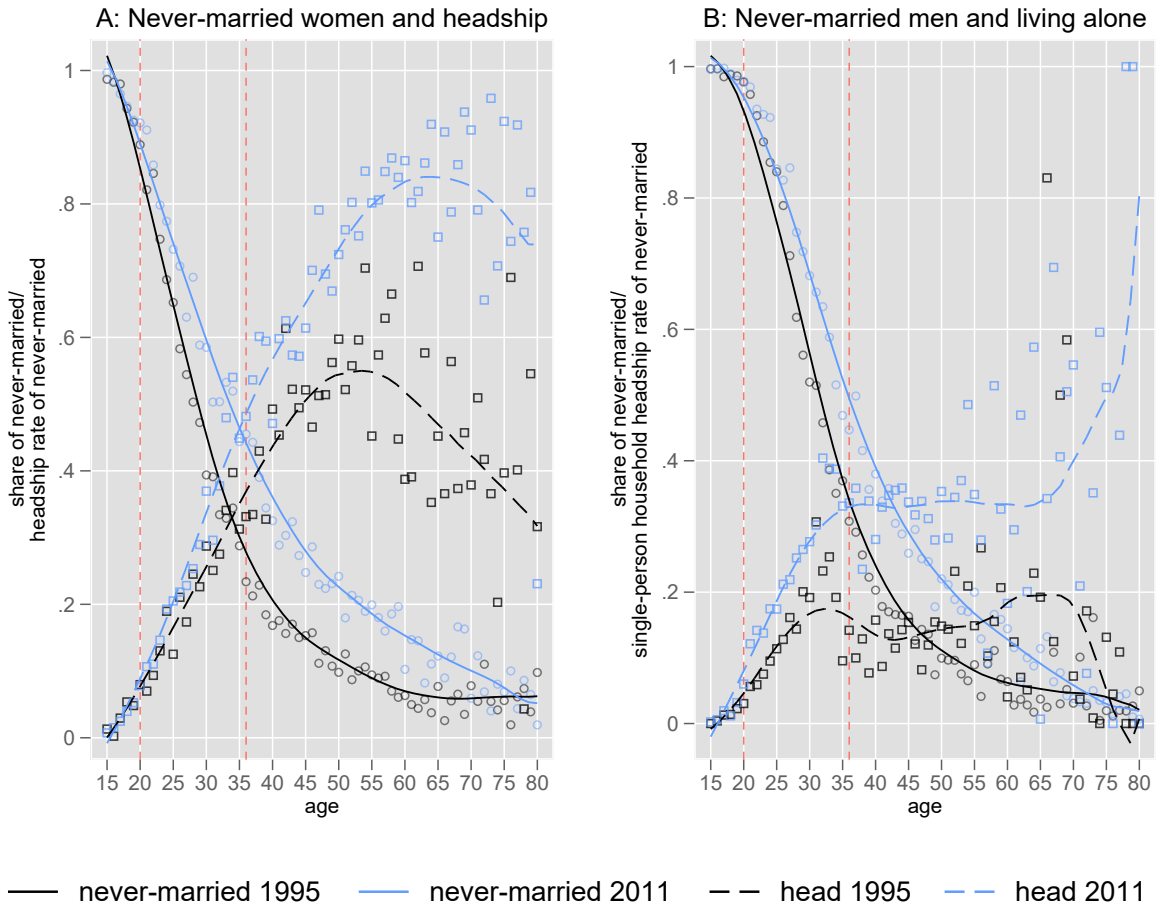
Figure 5.12: Headship rates conditional on marital and labour market status across age by gender in 1995-7 and 2009-11



Notes: own calculations using the OHS-GHS series and weighted using the Best CEW. Not Employed = unemployed or NEA. Unmarried = widowed, divorced/separated, or never married. Married = married or cohabiting.

A central story emerging from this section is that less marriage has implications for household formation. To show this link, we focus on never-married people in Figure 5.13 because this is the group that is growing most quickly and because people cannot return to this category over the life-cycle in the way they can switch between other marital statuses. We also focus on household headship as the main story for women and living alone as the main story for men. The figure plots the share of never-married people at the beginning and end of the period (solid line) and compares this to the rate at which never-married people form households (dashed line). Expansion in the headship rate of never-married people coincides with the same ages that face a longer duration of being never-married. In the red vertical line we have marked 20-year-olds who have comparable chances of being never-married in both 1995 and 2011 and whose household formation behaviour is not that different. But 16 years later, the 36-year-old in 2011 faces a much higher chance of still having never married and the headship rate rises accordingly. Crucially, we are seeing two changes in Figure 5.13: changes in conditions in the marriage market and changes in household formation behaviour. An interesting question which we turn to in the next chapter is which of these effects dominate in net terms of the change in the headship rate.

Figure 5.13: Declining marital prospects and faster household formation between 1995 and 2011



Notes: own calculations using the OHS-GHS series and adjusted using Best CEW weights. Smoothed using an Epanechnikov kernel-weighted local linear regression with a bandwidth of 5 years and adjusted using the Best CEW weights. Scatters in the respective colour (circles for the solid line and squares for the dashed line) are the share per age. Vertical red reference lines at ages 20 and 35.

5.6 Conclusion

In the sixteen years between 1995 and 2011, the household stock grew by 58%, but the population by only 28%. This expansion is owed to population growth, population aging, but also due to an increase in the rate at which people are forming households, especially single-person households. These changing patterns of household formation are highly gendered and relate to how the labour and marriage market in South Africa modulate men and women's means to support themselves. Key constituencies are women and Black African men who have each contributed towards increasing the headship rate at different points in time over the post-apartheid period. This manifested as a boom-plateau pattern in female headship where the boom happened amongst prime-aged and older women in the 1990s and the plateau in the decade of the 2000s, during which most new female household formation was due to adult population growth. Men on the other hand followed a slow-boil pattern in which headship rate change was mainly

level in the 90s but took off for Black African men in particular in the decade of the 90s, while other groups of men reigned in their household formation. Between 1995 and 2011, Black African men almost doubled the rate at which they formed single-person households. Key focal points are then middle-aged and older women in the years immediately after apartheid; and Black African men in the decade of the 2000s and especially in the case of living alone.

Black African men are very much the vanguard of the remarkable rise in single-person households over this period. At first this appeared to be mainly young men, but we observe that living alone has increasingly bled up the age distribution over time. We expect this change is related both to declining rates of marriage, but also to men's greater ability to live alone because they both earn more and are less responsible for childcare than women (Posel and Hall, forthcoming). Due to data constraints already discussed it is difficult to assess to what degree these single-person households may be housing migrants in stretched households. This is relevant because Black African men are the group historically associated with labour migration (Wilson, 2001). If we consider that the share of men living alone who were married (and potentially remitting to a family located elsewhere) declined from 26% in 1995 to 17% in 2011 while the share who were never-married increased from 58% to 71%, then potentially the archetype of the male migrant living alone could be waning in this period. This is not to say that labour migration and stretched households are less relevant - Bank et al. (2020) show this certainly would be an incorrect conclusion - but that ever more complicated dynamics related to the family, union formation, the labour market, and generational change are driving who lives alone.

Two narrative strands emerge over the chapter tying these changes together: the feminisation of the labour market and increased grant access for women; and the decline in rates of marriage and rise of the never-married head for both women and Black African men. The feminisation of the labour market in the late 1990s coincides with the boom period of female headship and likely also underpins the expansion of female age-specific headship rates that occurred at the same time. At the same time, female heads remain less likely to be employed than male heads and instead rely more heavily on grant income, access to which also expanded over this period. This profile coheres with prior work by Posel and Rogan (2012) investigating why female-headed households remain at a higher risk of falling into poverty than male-headed. A main mechanism identified by these authors is household compositional changes whereby female-headed households are more reliant on grant income and female labour market income, whilst women remain at a disadvantage in the labour market.

The decline in marriage and its implications for household formation are central to understanding household change in the post-apartheid period. Between 1995 and 2011, there was an extensive increase in the share of heads who were never-married. Never-married was the most common marital status for female heads by the end of the period, after being tied with widowhood in 1995. An important finding is that the *rate* at which never-married adults are forming households has clearly increased, and this appears to coincide with the concomitant decline in the rate of marriage. Marriage affects the household formation of men and women differently: being unmarried is strongly correlated with women forming households, but the opposite is true for men. However, in the case of living alone, being both not-married and employed is the key profile. This profile for single-person households means we can expect worsening marital prospects to potentially contribute to more single-person living and therefore more households in general. Our investigation of the role of marriage is continued in the next chapter.

Chapter 6

Sources of change in household headship and living alone

6.1 Introduction

The previous chapter used the reweighted data from Part I to describe trends in household formation between 1995 to 2011. We found that both the rate at which people headed households, as well as, the rate at which people lived alone grew by 5 percentage points, representing a 14% increase in the case of the former and a remarkable 83% increase in the rate of the latter. Our goal with this chapter is, firstly, to strengthen our ideas about the drivers of household formation developed over the course of Chapter 5 by considering them altogether within the setting of a model. Secondly, our goal is to then gain a better understanding of the sources of this growth by decomposing this model over time to identify in a descriptive way which model components contributed the most to faster household formation behaviour. Ingredients for our model are based in the trend analysis from Chapter 5, which related faster household formation to the feminisation of the labour market and expansion of grant income access in the 1990s; as well as, detecting an important role for the marriage market.

It is this role of the marriage market that is of special interest in this chapter. Chapter 5 described the rise of the never-married head: by 2011, never-married was the most common marital status for female household heads after being tied with widowhood in 1995. The decline in the marriage rate has resulted in a mushrooming share of never-married adults in the population at large; however, this is interestingly not the only factor behind the rise of the never-married head. Another part of the story emerging from Chapter 5 is the acceleration in the rate at which never-married people are forming households and living alone. This is further intertwined with how marital status interacts with gender and employment status. For women, being unmarried was the most important risk factor for heading a household regardless of employment status, but things were more complex for men. Married employed men were most likely to head households; but failing this combination, employed or married men faced a similar likelihood of

heading households. Living alone introduces its own variation; being unmarried and employed is the key risk profile regardless of sex. This complicated picture suggests there is more to be unpacked regarding the role of marriage and household formation.

The econometric literature on household formation in South Africa has focused mainly on the income driver of household formation, and less on marriage (Klasen and Woolard, 2009; Keller, 2004; Ebrahim et al., 2013). Most of this work is focused on answering questions about where South Africa's large unemployed population is living. The common finding is that unemployed people remain attached to the parental home, often supported by a pensioner, in the absence of State-sponsored unemployment support. If this is the case, the question of why South Africans would form more and smaller households in the face of persistent and rising unemployment, is even more provoking. Non-wage income has also been studied: the onset of pension income has been shown to strengthen a pensioner's say in decisions about who may live in the household (Ambler, 2016). A common pattern is for young children to move in, while working-age women move out when a household member starts receiving a pension (Ranchhod, 2017; Ardington et al., 2009; Edmonds et al., 2005). In other words, households recompose upon the onset of pension income, but do not necessarily shrink like they do in the developed world (Keller, 2004). Again, if this is the case, it might not be immediately obvious that pension income will be leading to smaller households.

Almost all of the econometric models of household composition in the South African literature exclude marital status as a control variable (Klasen and Woolard, 2009; Ebrahim et al., 2013; Ambler, 2016; Keller, 2004). This is done on the basis of not biasing estimates on their main variable of interest which is usually employment, and which is endogenous with marital status (Klasen and Woolard, 2009). This could prove to be a pertinent omission given that marriage has been linked to changing household form in the developed world. Delayed marriage entirely accounted for the rise in single-living in Japan between 1985 and 2010 where single-person households are today the most common household type (Raymo, 2015). Declining rates of marriage and rising rates of divorce have been linked to new family forms and more single-living in OECD countries between at least 1970 and 2008 (OECD, 2011); and poorer marital prospects were linked to more household formation by women in the United States (Craigie et al., 2018).

South Africa strongly coheres with developed country trends of shrinking household size, increased single-living and worsening marital prospects, but crucially diverges on other fronts. South Africa has much higher unemployment rates; a population comprised mainly of young people; much lower divorce rates; and historically high levels of circular labour migration, associated with single-living, compared to the West and other developed regions. In well-functioning labour markets where almost everyone is employed, it is understandable that household change could pivot on marital change. The link is less obvious in South Africa where unemployment is excessive. In a country where the labour market is at once critical for welfare and highly exclusive, one can imagine how gaining access to labour market earnings might be a priority that overrides marital status when it comes to household formation. This intuition is reflected in the weight of econometric research on household composition in South Africa focusing on income.

Why might marriage matter for household formation in South Africa's high-unemployment context? Firstly, marriage and labour markets are closely intertwined; and secondly, marriage has been the traditional way in which women access male earnings either because of female disadvantage in the labour market, or women specialising in household production (Lundberg, 2001). High unemployment will have

implications for both these points and, therefore, for household formation behaviour. As a clear illustration of this, Craigie et al. (2018) found in the United States that poorer local marital prospects (proxied by more unemployed men) were significantly associated with higher rates of headship by never-married women. This paper links household formation to marriage markets to labour markets, demonstrating one way in which they interrelate. Posel and Casale (2013) have already shown for South Africa that high levels of unemployment amongst Black African men in a local area undermines the chance that Black African women will be married. Potentially, a final link can be made here between marital prospects and household headship.

However, a focus on marriage for South Africa is also motivated by what has been observed in the literature on household composition. The post-apartheid period has witnessed an extensive increase in households composed of male or female adults only (Posel and Hall, forthcoming). Whilst declining rates of marriage have been somewhat offset by a rise in cohabitation in the West, in South Africa the same trend has resulted in men and women increasingly living apart from one another. This pattern has implications for livelihood strategies and welfare, as access to the labour market remains gendered and clusters at the household level (Pirouz, 2005; Wittenberg, 2017). Household compositional changes whereby female-headed households are increasingly reliant on female labour market earnings and grant income is identified as a chief mechanism behind their increased poverty risk compared to their male counterparts (Rogan, 2013; Posel and Rogan, 2012). These insights motivate the need to understand what might be driving these compositional changes, bringing the household formation process and marriage, as an institution governing gender relations, to the fore.

This chapter has two broad goals. First, we want to test the ideas flowing out of Chapter 5 more rigorously within the confines of a model, and find out which components have been most influential over time by decomposing our model. Secondly and within this overarching aim, we are specifically interested in how marriage fits into this process and through which channel the marriage effect has operated. The next section, then, considers how marriage matters for household formation from theory. Section 6.3 uses the OHS-GHS series to develop the research question about how the marriage market is influencing household formation, which is part of the motivation for our decomposition. To analyse how the model changes over time, we borrow from labour economics and apply an Oaxaca-Blinder decomposition to our model. This method descriptively picks up the most important sources of change in the model; as well as, to what extent each source operates via a change in conditions, a change in behaviour, or some combination of the two. The fact that our outcome variables are binary leads to a series of complications for the application of our model and decomposition that need careful consideration. We lay out these issues, motivate our choices, and describe our method in-depth in Section 6.4. Results for the model and the decomposition are in Sections 6.5 and 6.6, respectively. Acceleration in the rate at which never-married people form households emerges as one of our most important and robust results for both household headship and living alone. Marital status is found to strongly interact with labour market status emphasising once more that marriage and access to income are much more closely bound up together for women than for men.

6.2 Framework for studying household formation

The theoretical literature we are based in comes from Ermisch (1999) who compares indirect utilities of a young person moving out and forming their own household compared to remaining attached to their parental household. A simplification in this framework is that labour market status (and later, marital status) are taken as exogenous even though other theoretical (McElroy, 1985) and empirical (Keller, 2004) work has shown these decisions are often made endogenously. Klasen and Woolard (2009) argue that structural unemployment makes the assumption of exogenous labour market status defensible in this type of model for South Africa; although this is probably less the case for marriage. The basic components of our model are preferences for privacy and autonomy; the cost of moving out (housing prices); and the ability to afford to do so (wage and non-wage income) (Ermisch, 1999). It is assumed that more privacy and autonomy increases utility, meaning some privacy and autonomy cost is incurred by living with other people (Fafchamps and Quisumbing, 2007). Own access to wages or non-wage income increase the chance someone can form their own household, whilst higher house prices reduce this.

We extend this model for our purposes by considering the role of gender and marriage. Usually, preferences about privacy and autonomy are modelled as functions of age and education (Fafchamps and Quisumbing, 2007). Gender relations and the role of marriage in society, however, also modulate social norms about who may have how much privacy and autonomy, and when (Fafchamps and Quisumbing, 2007). We described in Chapter 2 that many South Africans subscribe to patriarchal ideas about gender. A stylistic description of patriarchal gender relations would be that men hold more authority and are more independent than women (Timol et al., 2019; Boonzaier, 2005; Albertyn, 2009). These norms make it more acceptable for young unmarried men to leave the parental home to 'establish a household' (an extension of autonomy). Women are expected to remain at home until they are married (and relatively sacrifice privacy and autonomy), upon which they join the household of either their husband or parent's-in-law, as wives (Goldscheider and DaVanzo, 1989; Siqwana-Ndulo, 1998). A married man is also more likely to form his own household to improve privacy for the married couple (Rosenzweig and Wolpin, 1993). Whereas the conventional pattern is that women default into headship upon the onset of widowhood or divorce - the loss of a husband is the socially acceptable time for a woman to take on more autonomy.[1] Acting against gender norms can be thought of as incurring some social cost, itself modulated by the extent to which gender norms clash with preferences about privacy and autonomy and the realities of conditions in the marriage market.

Patriarchal ideas about gender reinforce persistent norms about household gender specialisation, where women typically specialise in household production (housework and childcare) and men specialise in labour market production (Lundberg, 2001). This sets up female disadvantage in the labour market due to women's lost job experience from exits to care for children, as well as employers discriminating against women on the assumption childcare will make their tenure less stable (Lundberg, 2001; Folbre, 1995). This result has borne out in continued female disadvantage in the South African labour market, in terms of women both earning less and facing higher unemployment rates than men (Casale et al., forthcoming). The outcome is a perpetuation of women's dependence on men's earnings, traditionally accessed through marriage. Gaining a spouse usually means also gaining access to their labour market

[1] Many widows and divorcees are not allowed to remain in their husband's home in South Africa (Claassens and Ngubane, 2008) meaning it isn't clear that she would be gladly taking on more autonomy. Social norms could instead be insisting that she take on more autonomy and privacy whether she likes it or not.

earnings if we assume households pool income, which is the assumption underlying most welfare analyses using per capita household income. This makes marriage central to female livelihood strategies in a way that is not the case for men. In turn, this makes marriage at least as important a factor for women when it comes to the financial feasibility of forming a new household, as her own wage or non-wage income.

6.3 Setting the scene with survey data

The data we use is the same series of stacked cross-sectional household surveys used throughout the thesis, the OHS-GHS series weighted using the Best CEW. As such, we do not describe the data in detail but refer the reader to Part I where these data were discussed and set up. As was the case in Chapter 5, we conceptualise of household heads as the person most likely to have taken the step to form the household at some point in time and this makes them the subject of our analysis.[2] In keeping with the work from Chapter 5, we also model single-person households as a sub-population of special interest to post-apartheid household proliferation.

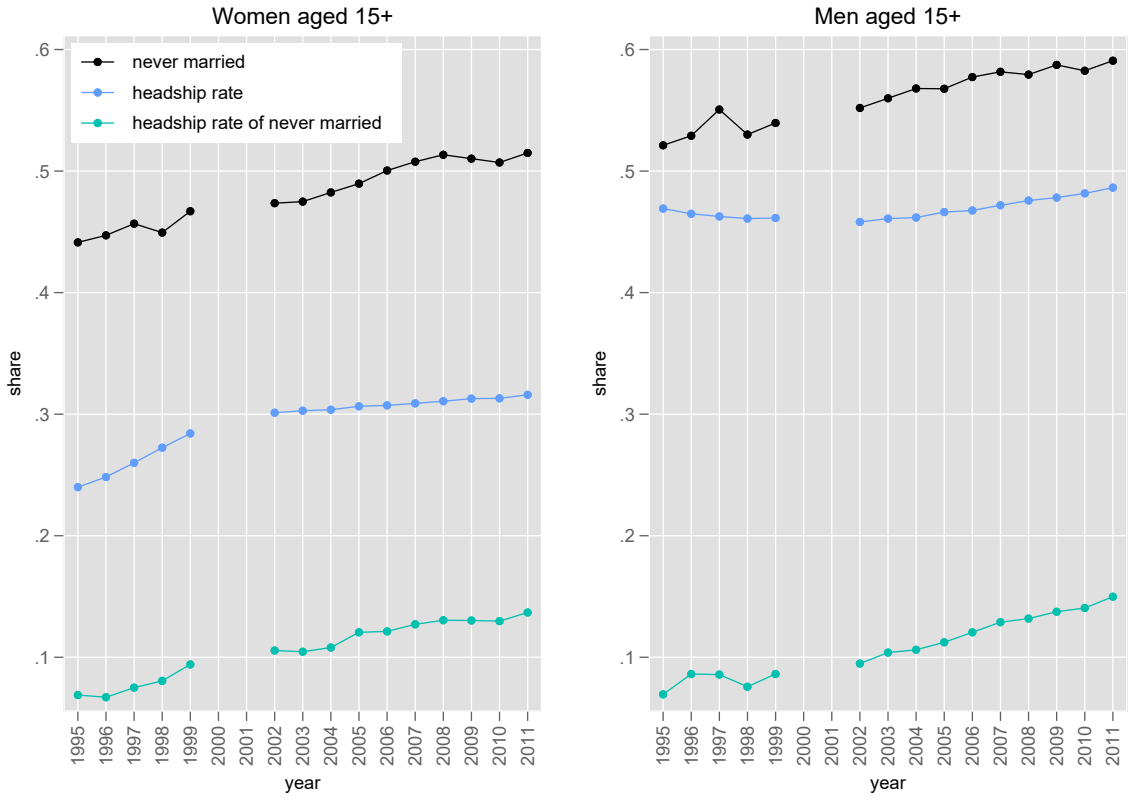
In this chapter, we are aiming to understand why household headship rates have increased over time. We describe this trend in our data in the blue line in Figure 6.1. We could show the same for single-person households, but use headship for now just to illustrate. The influence of changes in the marriage market on this change is of special interest in this chapter. Understanding its role, however, is complicated by the fact that we have observed extensive change both in the distribution of marital status in the population; as well as, the household formation behaviour of people with different marital statuses.

The share of never-married adults increased between 1995 and 2011, as captured by the black line in Figure 6.1. Based on the profiles from Chapter 5, this on its own could potentially explain why there were more single-person households and more women heading households, in general. But, we also found in Chapter 5 that the *rate* at which never-married people headed households increased, which could also be an important effect. This is captured by the green line in Figure 6.1. This leads to the question, are we seeing more households because there are more never-married people in the population, or because the household formation behaviour of the never-married population is accelerating? In other words, which has played a more important role: changes in composition or changes in behaviour.

The reason this question is interesting is because change owed more to behaviour or more to composition have different implications for our understanding of household formation. If the increased supply of never-married people could mainly explain more headship, then our understanding of how marriage influences household formation is quite mechanically linked to worsening marital prospects. For example, in Japan rising age at first marriage entirely accounted for the increase in single-person households (and not changes in how people of different marital statuses were distributed across different household types) (Raymo, 2015). This scenario is considerably simpler than if we find that change in household formation behaviour is more important. Changing behaviour opens up more questions about why this behaviour is changing and many influences could be at play. These include changing norms (e.g. women leaving the parental home before marriage), values and preferences (e.g. Individual vs. Family), how technology might ease and transform the meaning of solo living, and evolving livelihood strategies.

[2]Thinking of household heads as 'household formers' is an imperfect association. Chapter 5 already dealt with the merits and disadvantages of this approach. One of the main goals of this thesis is to explain household proliferation and the number of heads tally with the total number of households in our data.

Figure 6.1: Drivers of headship: behaviour and conditions related to marital status



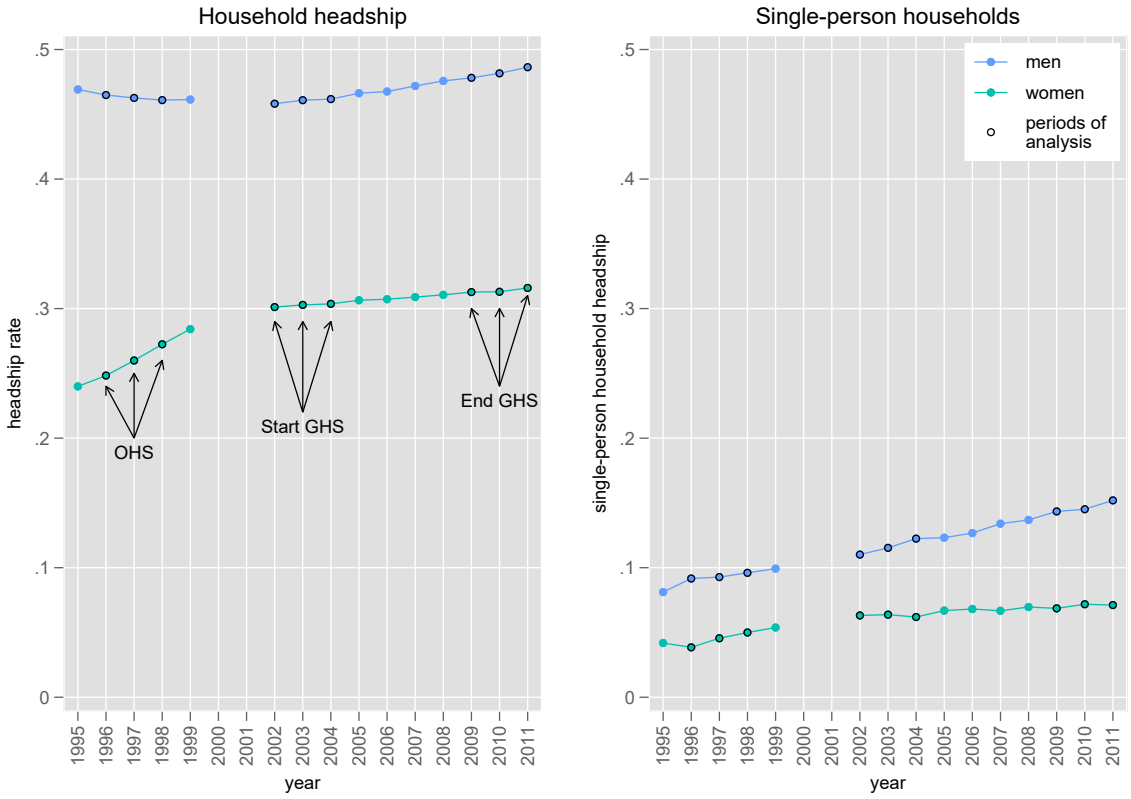
Source: own calculations using the OHS-GHS series weighted using the Best CEW.

To differentiate between these different influences we apply an Oaxaca-Blinder decomposition to our model over time. The decomposition serves to separate out the change in headship owed to behaviour versus attributes by running the thought experiment of holding one constant, changing the other, and computing how the headship rate would change. Figure 6.2 uses the OHS-GHS series data to plot what by now are familiar trends in rates of household headship and living alone, already reported in the previous chapter. Now, we identify periods of analysis between which to decompose changes in our model of household formation. In defining these periods of analysis, we pool a few years at a time because decompositions are sensitive in general to these types of choices.

Our Period 1 decomposition compares changes between samples pooled for years 1996-8 to those in years 2002-4, labelled 'OHS' and 'Start GHS' on the plot, respectively. We begin in 1996, and not 1995, because the level of employment in 1995 looks too high compared to the rest of the trend (see Figure 5.8). The Period 2 decomposition compares 2002-4 ('Start GHS') to 2009-11 ('End GHS'). We take this approach because the distinct trends between the OHS and the GHS for household headship in particular suggest its worth analysing these periods separately. At the same time, this acts as a robustness check, since the structural break in the patterns coincides with a change in survey instrument from the OHS to the GHS.

Rates of household headship and living a single-person household (SPHH) in each period of analysis

Figure 6.2: Periods of analysis for a decomposition of household headship and living alone



Source: own calculations using the OHS-GHS series weighted using the Best CEW.

are reported in Table 6.1, along with change in the rate between periods and sample sizes. Patterns of change reflect the boom-plateau pattern for female headship, and slow-boil pattern for male headship, described in Chapter 5. There is a slight decline of less than half a percentage point in the male headship rate between the OHS and Start GHS periods, but male headship increases by 2.45 percentage points in Period 2. The more important increase for women happens between the OHS and Start GHS periods, with a 4.15 point growth, followed by much more moderate growth in the GHS period of just 1.34 points. The rate at which men live alone increased at a similar rate in both Period 1 and Period 2, resulting in a 5.42 percentage point increase in total. Female single-person households followed the same pattern as their headship rates by growing more than twice as fast in Period 1. Overall, the more important change for men was in their rates of living alone; whereas, for women this was their expansion of household headship in general, and especially in the 1990s.

Table 6.1: Levels and change in rates of household headship and living alone in three periods of analysis with sample size

	Women			Men		
	OHS 1996-8	**Start GHS 2002-4**	**End GHS 2009-11**	**OHS 1996-8**	**Start GHS 2002-4**	**End GHS 2009-11**
Headship:						
Rate (%)	**26.07**	**30.22**	**31.56**	**46.41**	**45.94**	**48.39**
N	(105 296)	(109 657)	(105 269)	(83 997)	(93 468)	(87 360)
Change (pp):						
Period 1	4.15			-0.47		
Period 2		1.34			2.45	
Total	5.49			1.98		
SPHH:						
Rate (%)	**4.47**	**6.30**	**7.13**	**9.42**	**11.70**	**14.84**
N	(105 296)	(109 657)	(105 269)	(83 997)	(93 468)	(87 360)
Change (pp):						
Period 1	1.83			2.28		
Period 2		0.82			3.13	
Total	2.65			5.42		

Notes: own calculations using the OHS-GHS series weighted using the Best CEW. Household headship rate = share of household heads amongst the population aged 15 years and older. SPHH = single-person household. Single-person household rate = share of people living in one-person households amongst the population aged 15 years and older. The rate of single-person households is a subset of the rate of household heads in general. Sample limited to reflect regression model sample for reasons of having consistency throughout, meaning the sample is limited to observations for which all regression variables in Tables 6.3 and 6.4 are non-missing. This restriction alters the outcomes reported here very slightly at level of the first decimal point.

6.4 Method

There are two stages to our analysis, a model stage and a decomposition stage. The model is run for the OHS, Start GHS and End GHS periods and helps establish the relative importance of different drivers for a general model of either household headship or single-person living. We then decompose change in the outcome between these periods to assess which of these drivers contributed most towards increasing household formation and whether this was primarily via a compositional or behavioural channel. We describe the model in Section 6.4.4 and the decomposition in Section 6.4.5. Before we do so, there are some choices to be discussed relating to which covariates we include in our model (Section 6.4.1), model choice (Section 6.4.2), and complications with the decomposition (Section 6.4.3). It must be emphasised that our main result - that faster household formation by never-married people has contributed to higher rates of both household headship and living alone - is robust to all of these choices.

6.4.1 Scope

Data constraints mean that several factors our framework implies should be included in our econometric model, cannot be included. From our framework, we would want to include any variable that might impact preferences, the cost of forming a household, and the income available to pay for these costs.

Mainly data constraints but also the risk of overspecifying the model mean there are some key omissions, namely housing prices, fertility, labour migration, and the impact of the onset of democracy. Our data covers the period 1995 to 2011 and no sufficiently comparable data exist prior to the first democratic election in 1994. This means the extent to which changes in the 1990s are owed to the end of restrictions on where people could live is not something we can pick up. We also cannot properly trace fertility because our data only allows us to link mothers to children if they are living in the same household. Without a comprehensive fertility roster for each women we miss children who are living elsewhere or who passed away.[3]

We have already mentioned that we cannot identify labour migrants or labour-sending households in our data. This means the extent to which household proliferation has been the outcome of more or less labour migration (and more stretched households) is not something we can separate out. The best we can do is include variables that correlate with common characteristics of migrant labourer living arrangements, such as being urban and living in an informal dwelling (Posel and Marx, 2013). There is some evidence that labour migration was decreasing over the period we study (Posel, 2010) suggesting this is maybe not the main driver of household proliferation in our study, although the trend reversed after 2010 (Posel, 2020). An urban location dummy is included in our model which should control to some extent for the role of urbanisation. However, there are some quality shortcomings of this variable that should be noted which undermine the extent to which it is comparable over time. The definition of what areas counts as urban has changed over time (StatsSA, 2001*b*) and this, along with changes in the Master Sample of the GHS, has led to structural breaks in the series of the population share residing in an urban area.[4]

We are also unable to include housing prices in our model. There are two housing markets to consider, so to speak: the formal market and the market for free State-provided RDP housing. Sources for both of these prices are of very poor quality and so we do not include them in our model. In our data, a question asking "Did any household member receive a govt housing subsidy, such as an RDP housing subsidy?" can be consistently plotted over the GHS only[5] even though the State definitely was building houses before then, even if under the auspices of a different administrative programme (National Treasury, 2004). Survey questions are added and the module on housing subsidies becomes more sophisticated over time as government's own policy on housing became more complicated. However, we cannot extract a consistent series of questions over the full period or distinguish whether an affirmative answer to the above survey question pertains to an actual RDP house, a grant, or building materials. For example, the proportion of household heads answering affirmatively to this question went from 5.6 to 9.5% between 2002 and 2011. In a 2009 speech to parliament, the president stated that 2.6 million free houses had been built by the State (President Motlanthe, 2009); however, only 1.2 million households answered affirmatively to the question described above in 2009 (which covers more than just houses).

Outside of the survey it is surprisingly difficult to find out how many houses government has built. National Treasury's Intergovernmental Fiscal Reviews (IGFR) report the number of housing units completed and under construction as a combined number, instead of as separate figures. The delivery of State-built housing has been plagued by quality concerns and backlogs admitted in these reports, and further undermining the usefulness of these data (National Treasury, 2004). In any case, data is always

[3] Only the OHS surveys (1995-99) and the first survey of the GHS (2002) appear to ask questions about all live births.
[4] See Panel (b), Figure 3.4
[5] And OHS 1999, but we exclude this year from our sample

reported at the provincial level in which case province fixed effects included in our model should account for this. Essentially, the survey seems as good a data source as any others and it does not yield a series spanning the entire time period.

As for the private housing market, this data appears just as messy. The housing Consumer Price Index (CPI) is the best source of data because it goes back far enough in time. Privately curated housing data sources (e.g. Lightstone) seem to only begin in 2010 which is far too late for our purposes. Some of the banks curate their own housing price indices. ABSA's is the oldest but our inquiries were met with disinterest and the most useful information extracted from correspondence with ABSA was that this index has not been maintained recently and has undergone many changes. This bring us back to the CPI. The way that housing was measured changed fundamentally in 2008 meaning there is a break in the series in terms of measurement (StatsSA, 2008*c*). Additionally, historical housing CPI information is not available earlier than 2000 and, even after 2000, it is not available at a more disaggregated level than the whole country. In sum, house price data of any type is poor and we cannot properly account for it in our model.

Related to free housing is the free provision of services. Services can act as a sort of in-kind pricing of household public goods since the State provides free access to a base level of these services. Our data do include questions about household access to these various services. However, we exclude these from our model in an effort to keep the model parsimonious and avoid overspecification. Including these variables in our model did not change our results substantially.

Essentially, there are many factors to consider when modelling household formation given the complex nature of this social process. Accounting for all of these factors is beyond the scope of this work, partly because of data limitations. The focus of this chapter is on the interrelation between the marriage and labour market based on how both the local and international literature have highlighted these as key drivers of household change. Our effort is thus mainly one of incorporating the marriage market into a model of household formation that captures the effect of employment, grant income, and preferences. Given the existing econometric literature on household composition in South Africa, this represents a substantial extension on its own (Klasen and Woolard, 2009; Keller, 2004; Ebrahim et al., 2013). We then further extend this by running the Oaxaca-Blinder decomposition and including specifications for single-person households.

6.4.2 Model choice

Our outcome variables - household headship and single-person living - are both binary outcome variables. When modelling such variables, researchers usually choose between using a linear probability model (LPM) or a logit (or, probit) model, but the best choice is not obvious. The concern with using an LPM is twofold: heteroskedasticity in the errors and that estimates can exist outside of the unit interval, making them biased and inconsistent (Horrace and Oaxaca, 2006). Heteroskedasticity can easily be solved by using robust standard errors (which we do), but the latter concern is more of a problem. For this reason, many researchers prefer using a logit. Difficulty in achieving consistent results from an LPM does not imply a logit is then the correct probability model (Horrace and Oaxaca, 2006). Nevertheless, the logit is a popular modelling choice for binary outcomes and one consideration in the choice to use a logit over an LPM is how frequently outcome events occur (i.e. how frequently ones occur in your outcome variable

relative to zeroes). LPMs and logits will produce almost identical results when events occur around 50% of the time in the sample (Timoneda, 2020). The logit correction becomes more important as the chance of events occurring tends towards the extremes and the non-linear correction for the predicted probabilities becomes more important.

In the case of rare events, it has long been established in the statistical literature that the logit is biased, although this has been slow to filter into applied research (King and Zeng, 2001; Leitgöb, 2020). This bias is due to a multiplicity of problems including small sample size, low-dimensional data-sets, and models where the number of covariates is high relative to the sample size. However, King and Zeng (2001) show bias emerges even with good samples sizes in the thousands.[6] A recent paper by Timoneda (2020) considering the case of time-series cross-sectional data - which is our data format - argues that when including fixed effects, using an LPM with dummies is less biased than using either a logit with dummies or a conditional logit, especially in the case of rare events.[7] This naturally raises the question of what counts as a rare event. The usual rule of thumb is that an event occurring less than 5% of the time is rare, but under Timoneda (2020)'s conditions, the LPM outperformed the logit for events occurring less than 25% of the time. In our case, Table 6.1 shows we have some outcomes that occur close to half of the time (male headship) and others that could be considered rare (female single-person households in the OHS period). This discussion mainly serves to make the point that it is not categorically obvious that a logit is a superior modelling choice to the LPM for binary outcomes. An LPM will not provide the true marginal effect for a nonlinear model, but neither will the 'wrong' nonlinear model (Horrace and Oaxaca, 2006).

Model choice is further complicated by our goal to decompose the model. As we set up in Section 6.3, we are interested in decomposing the change in headship and living alone into changes owed to behaviour or composition. To do this, we apply the Oaxaca-Blinder (OB) decomposition (Oaxaca, 1973; Blinder, 1973), using Ben Jann (2008)'s *oaxaca* command in Stata 16. The OB method is often used in labour economics to study differences in outcomes (e.g. wages) between groups (e.g. sex, race, in our case, time) by setting up a counterfactual thought experiment using linear regression models (Jann, 2008). There are two versions of the OB method, a twofold and threefold.[8] The twofold is by far the more popular of the two in economics (Biewen, 2012). It decomposes the group differences into an 'explained' and 'unexplained' portion and the latter is often used as a measure of discrimination. The threefold version that we use here decomposes differences in the outcome between groups into an 'endowment', 'coefficient', and 'interaction' effect. We motivate our choice of the threefold over the twofold at the end of Section 6.4.6, since it makes more sense coming after our description of the decomposition in Section 6.4.5.

Economists have long been able to extend the OB to the non-linear case. In the linear case, counterfactual expectations are computed by applying the predictor means from one group to the coefficients from another.[9] In the non-linear case, this simple procedure is complicated by the non-linear link function causing the conditional expectations to depend on the values of all other predictors and not just the

[6] Evidence is mixed about the direction of this bias with some authors finding it leads to smaller predicted probabilities (King and Zeng, 2001) and others finding the opposite (Timoneda, 2020; van der Paal, 2014).

[7] Part of why this is the case is because maximum-likelihood estimation will drop groups without variation resulting in a form of selection on the dependent variable; whereas this is not the case with the LPM which will allocate a zero average to that group. We have time-series cross-sectional data and include province fixed effects, but our sample size of more than 500 000 is many times more the number of groups, which is nine.

[8] Winsborough and Dickinson (1971) originally describe a fourfold version from which the three- and twofold are derived and best practise today would be either the twofold or threefold.

[9] Details in Section 6.4.5

means (Fortin et al., 2011). Different methods have been proposed to deal with this issue which have been distributed in different commands for Stata (Fairlie, 2005; Powers et al., 2011; Bauer and Sinning, 2008; Bartus, 2006) (although none apply the threefold decomposition except Jann (2008)'s *oaxaca*). However, extracting detailed decomposition output for non-linear models is more complicated than in the linear case. Detailed decomposition output refers to the coefficient, endowment and interaction effects disaggregated by individual covariates. For example, we are interested in the coefficient effect of marital status on headship, not the coefficient effect of all our covariates combined. Extracting these effects in the case of a linear model is simple because the linearity means the model is additively separable and the sum of all effects will equal the gap in the outcome (Jann, 2008). This is not so simple for non-linear cases because non-linearity of expected values (and therefore, the gap) means the contribution of a particular effect depends on the values of all other predictors. A number of solutions have been proposed but according to Jann (2018), there is no best way of dealing with this.

The *gdecomp* command from Bartus (2006) uses the average of the marginal effects and then applies the standard OB. The problem in this case is that the sum of contributions from different covariates and effects may not add up to the total gap. Fairlie (2005) provides a command called *fairlie* which sequentially switches the covariates to create a series of counterfactuals and then reports the average of this process. This process is complicated when the groups are not the same size or when using sampling weights leading to an additional matching problem. The bigger problem, though, is that the results are path dependent. The suggested solution is to randomise the order in which covariates are added to the decomposition and then average over a certain amount of repetitions.

The method that has gained the most traction in the literature is a linearisation technique from Yun (2004) that is invariant to the order in which covariates enter the decomposition, thus providing a solution to path dependency. This method is provided in the Stata command *mvdcmp* (Powers et al., 2011) but is also conveniently incorporated into Jann's *oaxaca* command when specifying the logit or probit option. It works by first approximating the gap between groups at the means of covariates and then linearising the non-linear regression models using a first order Taylor expansion. After linearising, weights for each component are constructed using the contribution of each to the linearisation process, so that detailed decomposition output can be extracted and so that they sum to the correct total. The issue with this approach is that there is no indication of the quality of the linearisation. As Jann (2018) points out, if differences in either coefficients or (the means of) covariates are large, then the approximation could be poor. Perhaps more importantly, if most of the data is in a highly nonlinear region of the logit or probit function, as might be the case with rare events like single-person households, the approximation may also be poor.

Jann (2018)'s suggestion for binary outcome variables, therefore, is to use an LPM and run the standard OB. The LPM models conditional probabilities making it easily interpretable both for the regression output as well as the decomposition output; and it is straightforward to extract detailed decomposition output. Jann (2018) points out that it is not obvious from first principles that the Yun (2004) decomposition on a logit or probit would be better than an OB on an LPM and that both approaches will have similar problems if the linear approximation is poor. The trade-off we face then is between the simplifying functional form assumptions of the LPM and the quality of the linearisation for the Yun (2004) decomposition. The former approach is potentially biased for our rarer outcome, single-person households, but sturdy for our more common outcome, household heads, and we can trust

the detailed decomposition output is precise. In the latter approach, the logit model might be less biased for single-person households (although this is not a certainty based on the above discussion) and the extraction of detailed decomposition output might be noisier. Our approach is to follow Jann (2018)'s recommendation and maintain the LPM and the OB as our main model in the text, but run a logit and apply the Yun (2004) decomposition in the appendix. The Yun (2004) decomposition is then laid out in more detail in the Appendix B.4. Ultimately, our results are highly robust.

6.4.3 Base category problem

Another well-known problem that afflicts all versions of the decomposition is the base category problem (Fortin et al., 2011; Oaxaca and Ransom, 1999; Jones and Kelley, 1984; Gardeazabal and Ugidos, 2004). When using categorical variables, researchers often include sets of dummies and exclude one which becomes the base category relative to which coefficients on the included dummies are interpreted. However, changing which category is omitted, changes the slope coefficients and the intercept. The effect is to change the decomposition results, and specifically, the coefficient effects (or, unexplained effects in the case of the twofold decomposition) both in terms of the single dummies; as well as, the contribution of the categorical variable as a whole (but not the aggregate coefficient effect) (Jann, 2008). It is impossible to tell which part of the coefficient effect is truly owed to group membership (the intercept) versus owed to differences in the effect of the omitted category (Fortin et al., 2011). Often the choice of which category to omit is arbitrary or based on practical modelling factors and, in fact, this problem can extend to continuous variables which have no meaningful interpretation of the zero-value or which have been scaled, e.g. standardised test scores (Jones and Kelley, 1984).

Suggestions have been made to mitigate this problem, but Fortin et al. (2011) emphasise that there is no solution and the problem should be viewed as a conceptual one. Gardeazabal and Ugidos (2004) and Yun (2005) suggest different ways of neutralising the effect of the base category on the intercept by normalising the categorical variable and forcing the coefficients to be expressed as deviations from the grand mean. Yun (2005)'s solution is to restrict the coefficient on the base category to reflect the unweighted average of the coefficients on the other categories, although admits some arbitrariness still exists in deriving this normalisation equation.

Fortin et al. (2011) argue that the problem with normalisation is that it can leave the decomposition without a meaningful interpretation and might preclude comparisons across studies because results are sample specific. Instead it is recommended to choose a meaningful base category. In literature comparing returns to schooling, for example, Fortin et al. (2011) point out that it is common to make high school graduates the omitted category, meaning this practise is well-established and makes comparison across studies accessible. They argue against trying to mimick continuous variables by making the least well-remunerated category ('less than high school') the base, because this category may vary more across studies.

This may be a strong argument for the returns to schooling literature, but the econometric literature on household headship and single-person models is much less established and there is no norm in the literature for us to turn to. However, the point to choose a meaningful base category is well-taken and this was what we aimed to do. There are two considerations in this case: mainly marital status, and secondly, age. The two most sensible marital statuses to use as base categories are 'married' or 'never-

married'. Over the course of Chapter 5, we made a case for an important role of never-married people and so we want this category to appear in our output for accessible interpretation. This means 'married' is our omitted category.

This has implications for age. It seemed sensible at first to include age as a quadratic (i.e. age and its square). However, this meant that our base category was then married 15-year-olds, which does not feel like a sensible base category in the spirit of Fortin et al. (2011). This motivated us to instead include age as a categorical variable, with categories being 15-24 years, 25-34 years, 35-44 years, 45-59 years, and 60 years and older. Doing so allows us to use married 35-44-year-olds as our base; middle-aged married people feels more like a meaningful category. Some downsides of this choice are that we are potentially exacerbating the base category problem in our results by including a variable in categorical form when it need not be. However, having 'never-married' reported in our output seemed more important, and an advantage of this approach is a much more interpretable age effect that can be interacted with pension access.

Our base category is then married Black Africans aged 35-44 years-old who are not employed and have incomplete secondary schooling living in an urban area of the Western Cape in formal housing and not receiving a pension (by definition). More detailed information about the model and other variables is in the next section. We opted to use 35-44-year-olds as the base because the constant term in the LPM was most stable using this age group. More importantly though, in the next chapter 15-34-year-olds count as our definition of young adults and it may be worth not splitting them up (i.e. not using the other obvious candidate of 25-34-year-olds as the base) so young adults as a whole can be interpreted in the output. All other bases were chosen based on the most common category so that our base could be described as highly representative of the average South African.

6.4.4 The headship and single-person household model

We run an LPM in which indicator variables for headship (H) and single-person living (S) are regressed on sets of explanatory variables. The regression is weighted using the Best CEW series and we use cluster-robust standard errors to adjust for heteroskedasticity in the errors.[10] The regression is run separately for sex for each of our three time periods (OHS, Start GHS, End GHS) and limited to adults aged 15 years and older. Marital status is incorporated into our model by interacting it with labour market status based on our theoretical framework and previous descriptive work. A categorical marital status variable, MS, is interacted with employment status, E.[11]

Other covariates in the vector X are based on the framework discussed above, with variable definitions in Table 6.2. The main preference function variables are age and education. We also control for own pension receipt as a form of non-wage income, but only people over the age of 60 are eligible for pensions so this variable enters as an interaction effect with the top age category of people aged 60 years and

[10]The clusters in all years, with the exception of 1998, are the enumeration area, or primary sampling unit, used by StatsSA which usually comprise 100-250 households and from which a sample of households are drawn to be surveyed. There are about 2800-3000 clusters in each year of the GHS. In the OHS period, there are 747 clusters in 1996 (which had a smaller sample than other years) and 1049 cluster in 1997. In 1998, the enumeration area variable could not be found in the data set, so we used the next smallest geographical region available in the data which was local municipality district, of which there were 319 in our sample.

[11]For reasons of space, we immediately present a model including an interaction in the main text. Readers interested in how the regression results change when marital status, employment status, and their interaction are added iteratively are referred to Appendix B.1.

Table 6.2: Definitions of regression variables and vectors

Variable	Definition
H	1 = household head; 0 = not a household head
S	1 = single-person household; 0 = multi-person household
MS	categorical: married (base), never married, divorced/separated, widowed
E	1 = employed; 0 = not employed (NEA or unemployed)
P	categorical: 9 provinces with Western Cape as the base
X	
age	categorical: 15-24 years; 25-34 years; 35-44 years (base); 45-59 years; 60+ years
edu	categorical: primary or less; incomplete high school (base); high school/matric; post-secondary
pension recipient	1 = pension recipient and over 60 years; 0 = not pension recipient or under 60 years
rural	1 = rural location; 0 = urban location
informal dwelling	1 = informal dwelling (e.g. shack, backyard dwelling, caravan tent); 0 = formal dwelling (e.g. house, flat, traditional dwelling)
race	categorical: african (base), coloured, asian/indian, white

more.[12] Some spatial characteristics are also included in X in the form of a rural location dummy and whether a respondent is living in an informal dwelling. These variables along with province fixed effects (P) will hopefully capture some of the variation in access to free housing which is administered at the provincial level. Finally, we also motivate the inclusion of race as a critical proxy for legacy effects of apartheid that explain variation in education, labour market outcomes, and marital prospects.

For outcomes $Y = \{H, S\}$, the following specification is run for each individual i in a province p:

$$Y_{ip} = \beta_0 + \beta_1 \sum_{i=1}^{n} MS_{ip} + \beta_2 E_{ip} + \beta_3 \sum_{i=1}^{n} MS_{ip} * E_{ip} + \beta_4 X_{ip} + \beta_5 P_p + \epsilon_{ip} \tag{6.1}$$

6.4.5 Oaxaca-Blinder decomposition method

As introduced in Section 6.4.2, the OB decomposes the differential in an outcome between two groups by setting up a counterfactual thought experiment using linear regression models. Here, we lay out the method in detail; following which, we motivate our choice of the threefold over the twofold version of the OB. Following Jann (2008), consider the difference (D) in the headship rate (H) over two periods of time, t_0 and t_1 (we use the headship example to describe the procedure, but the same applies to the single-person household case, S):

$$D = E(H_1) - E(H_0) \tag{6.2}$$

[12] We did not include access to the Child Support Grant since this would imply information about other household members and be endogenous to household formation.

where H can be modeled using a linear model:

$$H_t = X_t\beta_t + \epsilon_t, \tag{6.3}$$

$$E(\epsilon) = 0$$
$$t \in (0, 1)$$

The difference, D, can then be expressed as a difference in the linear prediction at the time-specific means of the explanatory variables:

$$D = E(H_1) - E(H_0) = E(X_1)'\beta_1 - E(X_0)'\beta_0 \tag{6.4}$$

This expression can be further rearranged in terms of contributions from differences in explanatory variables, coefficients, and both at the same time:

$$D = [E(X_1) - E(X_0)]'\beta_0 + E(X_0)'(\beta_1 - \beta_0) + [E(X_1) - E(X_0)]'(\beta_1 - \beta_0) \tag{6.5}$$

where the terms correspond with the endowment (W), coefficient (C), and interaction effect (I):

$$D = W + C + I \tag{6.6}$$

The W component is then the expected change in headship in t_0 if people in t_0 had the predictors of t_1; the C component is the expected change in headship in t_0 if people in t_0 faced the coefficients of t_1. The I component relates to the change in headship that cannot accurately be allocated to either purely a change in predictors or a change in coefficients, but instead the simultaneous changing of the two. We illustrate the decomposition in Panel A of Figure 6.3, following a similar example from Biewen (2012). The notation is simplified so that $E(X_t) = \overline{X_t}$ and $E(H_t) = \beta_t\overline{X_t}$. The decomposition serves to ask the question 'what must change in t_0 (behaviour or conditions) to get the higher headship rate of t_1?'. For household formation, we interpret the endowment effect as capturing change owed to changing conditions (e.g. more never-married people); and the coefficient effect as capturing behavioural change (e.g. differences in the rate at which never-married people formed households).

This is a good moment to emphasise that the decomposition remains a descriptive exercise. One might expect behavioural changes to interact with compositional changes, not in the sense of the interaction term, but in a causal sense. For example, the fact that fewer people are getting married might prompt never-married people to form their own households sooner and not wait around to make a match first. As more never-married women start heading their own households or living alone, this might serve to further normalise this outcome with potential knock-on effects for social norms. The point is that compositional change itself could cause behavioural change, but the OB decomposition cannot disentangle these complicated effects and measures only the net effect.

6.4.6 Motivating the threefold decomposition over the twofold

As mentioned above, the twofold decomposition is substantially more popular in economics (Biewen, 2012; Fortin et al., 2011; Jann, 2008) making it worthwhile justifying why it is not our preferred approach. The threefold decomposition is different to a twofold in its treatment of the I component. Where the threefold maintains I as a separate term, I is combined into either the W component or the C component in the twofold decomposition depending on how the decomposition is set up. In discrimination studies, researchers are often interested in decomposing a wage gap between two groups into effects owed to characteristics, called the explained effect, and effects owed to coefficients, called the unexplained effect and interpreted as a measure of discrimination. Consider the case of the gender wage gap where, to fit in with our previous notation, D is now the difference in wages, men are group 1, and women are group 0. The researcher faces a conceptual choice to either treat men as paid too much and evaluate the decomposition using women's wage level; or, to treat women as underpaid, and evaluate the decomposition at men's wage level. The difference is what Jones and Kelley (1984) call the 'privilege' and 'deprivation' models, respectively.

In the case of the privilege model, the C component in Equation 6.5 is untouched, but the W and I components are combined by multiplying out and rearranging terms to get the following:

$$D = [E(X_1) - E(X_0)]'\beta_1 + E(X_0)'(\beta_1 - \beta_0) \tag{6.7}$$

The first term then asks how men's wages would change if their characteristics were made to look like women's; the second term asks how much more women would be paid if they were paid like men (Jones and Kelley, 1984). Running the deprivation model runs the opposite thought experiment, this time keeping W intact, and yields:

$$D = [E(X_1) - E(X_0)]'\beta_0 + E(X_1)'(\beta_1 - \beta_0) \tag{6.8}$$

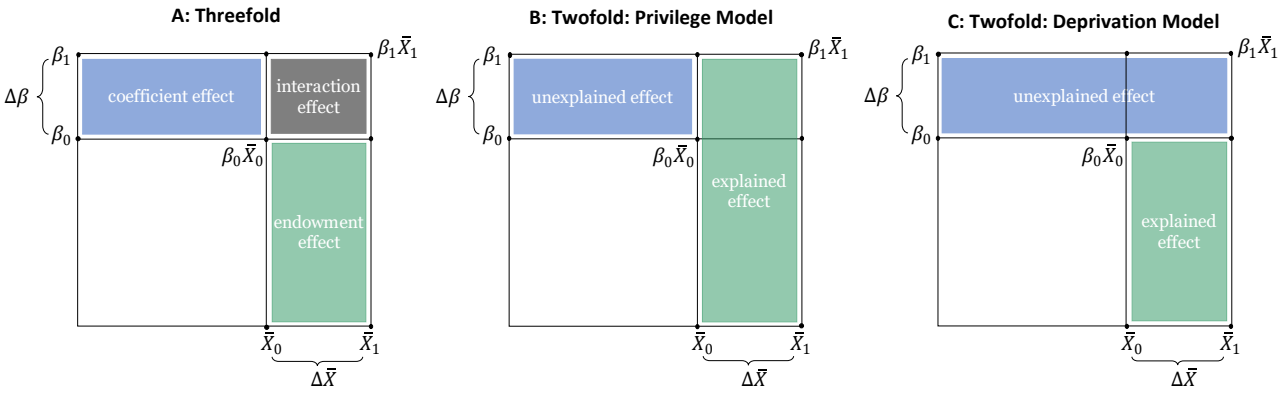
As such, the interaction effect is subsumed into the endowment effect in the privilege model to create a larger explained effect; and into the coefficient effect in the deprivation model to create a larger unexplained effect. This is illustrated Panels B and C in Figure 6.3.

Either of these set-ups could be justified depending on the mental experiment being carried out and it is up to the researcher to present a conceptual argument motivating their chosen approach. However, many researchers feel that neither approach is satisfactory since privilege for one group implies deprivation for the other. In many cases the wage structure of each is instead compared to an ideal non-discriminatory wage structure, called β^* (Jann, 2008). This introduces a new problem because β^* is not observed and so needs to be modelled and estimated. A common solution is to use the coefficients from a pooled regression of men and women adjusted by a relative weight given to each group (Neumark, 1988; Oaxaca and Ransom, 1994). This approach will still result in some portion of the unexplained component being allocated to the explained (Fortin, 2006; Jann, 2008).

The main point for our purposes is that a portion of the gap that usually cannot unambiguously be allocated to either stemming from change in coefficients only or endowments only, is subsumed into one of these effects. This occurs because of the mental experiment researchers of discrimination usually want to run. But this is not the same experiment we want to run. There is no reason why we would conceive

of household formation patterns in either t_0 or t_1 as ideal, or posit some unobserved ideal version of household formation that would be captured by β^*. Our question is more neutral in simply wanting to describe the change. There is therefore no need for us to muddy either our endowment or coefficient term by mixing the interaction term into one of them. Moreover, there is a conceptually coherent interpretation to the interaction term in our case: change in household headship owed to the simultaneous change in behaviour and conditions is meaningful. This might not be the case for discrimination studies.

Figure 6.3: Why is $\beta_1\overline{X_1}$ greater than $\beta_0\overline{X_0}$? Different versions of decompositions by group



Notes: adapted from Biewen (2012)'s Figure 1 illustration of the Winsborough and Dickinson (1971) decomposition and using Jones and Kelley (1984)'s naming conventions. $\overline{X_t}$ are the predictor means for group t; β_t are the coefficients for group t.

6.5 Regression results: correlates of household formation

The output for the regression on headship is in Table 6.3 and single-person household headship is in Table 6.4. The respective output for the logit models run as robustness checks are in Appendix Table B.2 and Appendix Table B.3, respectively. The results of the linear and logit models are highly comparable. The constant term in the LPM output reflects the chance that our base category either heads a household or lives alone. The base category is married 35-44-year-olds meaning that for men the chance of heading a household in Table 6.3 is very high (around 70%), but lower for women (declining from 25% in the OHS period to 18% in the End GHS period). Similarly, it is rare that married 35-44-year-olds would be living alone meaning the base probability is about zero for the specifications in Table 6.4.

There are four main points we want to draw out of this output. The first point is that marriage has opposite effects on male and female headship, but works in the same direction for living alone. In Table 6.3, marriage makes men more likely to head households, but the opposite is true for women. Never-married women were 18 percentage points *more* likely than married women to head households in the End GHS period, whereas never-married men were 40 percentage points *less* likely than their married counterparts in the same period. In Table 6.4, though, being never-married makes both men and women more likely to live alone. By the End GHS period, never-married women were 6 percentage

Table 6.3: Regression results for the headship model

YVAR: Headship	Women			Men		
	OHS 1996-8	Start GHS 2002-4	End GHS 2009-11	OHS 1996-8	Start GHS 2002-4	End GHS 2009-11
MS: never married	-0.00	0.10***	0.18***	-0.51***	-0.46***	-0.40***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
MS: divorced/separated	0.38***	0.42***	0.49***	-0.30***	-0.32***	-0.22***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
MS: widowed	0.47***	0.57***	0.58***	-0.20***	-0.16***	-0.05**
	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)
E: employed	-0.02**	0.05***	0.03***	0.11***	0.15***	0.13***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
MS*E: never married	0.28***	0.28***	0.21***	0.11***	0.11***	0.13***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
MS*E: divorced/separated	0.19***	0.18***	0.12***	0.12**	0.12***	0.09**
	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)
MS*E: widowed	0.17***	0.07***	0.06***	0.10*	0.10**	0.05*
	(0.01)	(0.01)	(0.01)	(0.04)	(0.03)	(0.03)
age: 15-24 years	-0.20***	-0.26***	-0.32***	-0.19***	-0.23***	-0.26***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
age: 25-34 years	-0.09***	-0.12***	-0.15***	-0.06***	-0.08***	-0.12***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
age: 45-59 years	0.03***	0.05***	0.08***	0.07***	0.06***	0.10***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
age: 60+ years	-0.01	0.02	0.07***	0.11***	0.14***	0.16***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
pension recipient (aged 60+)	0.03**	0.04**	0.04**	0.08***	0.05**	0.09***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
edu: primary school or less	0.01*	0.02***	0.03***	-0.00	-0.00	-0.00
	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
edu: high school/matric	-0.00	-0.01*	-0.02**	0.03***	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
edu: post-secondary	0.02*	0.04***	0.06***	0.09***	0.09***	0.11***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
race: coloured	-0.11***	-0.12***	-0.10***	-0.04***	-0.07***	-0.13***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
race: asian/indian	-0.14***	-0.15***	-0.14***	-0.02	-0.07***	-0.13***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
race: white	-0.11***	-0.12***	-0.11***	-0.00	-0.09***	-0.09***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
informal dwelling	0.08***	0.11***	0.16***	0.12***	0.15***	0.22***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
rural	0.01	-0.01	-0.00	0.04***	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Province FE	YES	YES	YES	YES	YES	YES
_cons	0.25***	0.20***	0.18***	0.71***	0.69***	0.69***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	105 297	109 657	105 269	83 998	93 468	87 360

Notes: own calculations using the OHS-GHS series weighted using the Best CEW on a sample of respondents aged 15 and older. * p<0.05, ** p<0.01, *** p<0.001. Cluster-robust standard errors in parenthesis. Base groups for categorical variables are as follows: MS = married; MS*E = married or not employed; age = 35-44 years; edu = incomplete secondary schooling; race = Black African; prov = Western Cape.

points more likely and never-married men were 15 percentage points more likely to live alone than their married counterparts. Being widowed or divorced makes people most likely to live alone. Being widowed is also one of the most important correlates of female headship: widows were 58 percentage points more likely to head their own households in the End GHS period compared to their married counterparts.

Being both employed and unmarried increases the chance that people will form households or live alone: the interaction effect was highly significant and positive in almost all specifications for both sexes. This relates to our second important result, which is that the interaction effect is more important for women than for men, and appears to soak up most of the effect of employment for women because the effect of being employed on its own is much more muted for women than for men. Consider that employment alone increases the chance that a woman will head a household in the End GHS period by just 3 percentage points, all else equal, compared to 13 percentage points for men. By contrast the combination of being never-married and employed increases her chance of headship by an additional 21 percentage points in the same period, compared to only 13 percentage points for men. This underlines that being both employed and unmarried is a particularly important route to female household formation; whereas, there is more diversity for men who might leverage either employment, marriage, or both to form a household. This finding is slightly different for male single-person households where the interaction is relatively more important than in the male headship model. This means that for both sexes, being both employed and unmarried is an important risk factor for living alone.

This brings us to the third point which is that the marital status effects are especially strong for men, potentially linked to the strength of patriarchal gender norms. These effects are interestingly larger than the employment effects. For example, the headship model predicts that a never-married employed man is still 14 percentage points less likely to head a household than a married not-employed man, all else equal. The strong marital status effect overpowers the positive effects of both employment on its own and its interaction with being never-married. The marital status effect also emerges as more important in men's single-person model. Just as being married was a key determinant of men heading households; being never-married is the key determinant for men living alone, with an important interaction with employment.

The fourth key point is that over time, the influence of the never-married term on its own strengthens. The never-married coefficient increases from zero to 18 percentage points for women between the OHS and End GHS periods. Never-married men increase the chance they would head a household relative to married men by 9 percentage points over time. Divorced and widowed men and women have also become increasingly likely to head households over time relative to their married counterparts. Similarly for living alone, the chance a never-married man would live alone trebled between the OHS and End GHS period. The size of the employment effect is more stable by comparison across specifications. The exception here is female headship between the OHS and Start GHS periods, where the employment coefficient changed sign and grew by 7 percentage points. Indeed, the coefficient on employment for women in the OHS period in Table 6.3 is in fact slightly negative.

There are some interesting results from the other covariates. The results reflect that headship increases at a decreasing rate over age, reflecting the age profile from Figure 5.3 in the previous chapter. Similarly, the cubic age profile for women living alone shown in Figure 5.6, also comes out clearly in the regression results. The results for headship show the same change across time for both sexes, which is that older people are becoming more likely to head households. By the end of the period, there were notable increases

Table 6.4: Regression results for the single-person household model

YVAR: Single-person household headship	Women			Men		
	OHS 1996-8	**Start GHS 2002-4**	**End GHS 2009-11**	**OHS 1996-8**	**Start GHS 2002-4**	**End GHS 2009-11**
MS: never married	0.04***	0.05***	0.06***	0.05***	0.08***	0.15***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
MS: divorced/separated	0.09***	0.09***	0.12***	0.25***	0.21***	0.33***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
MS: widowed	0.07***	0.08***	0.10***	0.18***	0.15***	0.26***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)
E: employed	0.00	0.02***	0.01	0.06***	0.08***	0.06***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
MS*E: never married	0.12***	0.18***	0.15***	0.11***	0.12***	0.15***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
MS*E: divorced/separated	0.09***	0.09***	0.06**	0.11**	0.10***	0.09**
	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)
MS*E: widowed	0.03*	0.03**	0.03*	-0.04	0.08**	0.07*
	(0.01)	(0.01)	(0.01)	(0.04)	(0.03)	(0.03)
age: 15-24 years	-0.03***	-0.03***	-0.05***	-0.06***	-0.11***	-0.15***
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
age: 25-34 years	-0.00	0.01*	-0.01	0.02**	-0.01	-0.05***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
age: 45-59 years	-0.01*	0.00	-0.01**	0.00	-0.01	0.01*
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
age: 60+ years	0.02***	0.07***	0.08***	0.01	-0.01	0.02*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
pension recipient (aged 60+)	-0.04***	-0.08***	-0.07***	-0.01	-0.00	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
edu: primary school or less	-0.00	0.00	-0.02***	0.01**	0.01	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
edu: high school/matric	0.01	-0.00	0.00	0.00	-0.02***	-0.01*
	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
edu: post-secondary	0.02**	0.02*	0.06***	0.02*	0.01	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
race: coloured	-0.02***	-0.02***	-0.04***	-0.04***	-0.07***	-0.08***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
race: asian/indian	-0.01	-0.01	-0.00	-0.02*	-0.06***	-0.08***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
race: white	0.05***	0.07***	0.06***	-0.02	-0.03***	-0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
informal dwelling	0.06***	0.09***	0.12***	0.16***	0.23***	0.23***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
rural	-0.01	-0.01**	-0.01*	0.04***	0.02*	-0.00
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Province FE	YES	YES	YES	YES	YES	YES
_cons	-0.00	-0.03**	-0.00	-0.04*	-0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	105 297	109 657	105 269	83 998	93 468	87 360

Notes: own calculations using the OHS-GHS series weighted using the Best CEW on a sample of respondents aged 15 and older. * p<0.05, ** p<0.01, *** p<0.001. Cluster-robust standard errors in parenthesis. Base groups for categorical variables are as follows: MS = married; MS*E = married or not employed; age = 35-44 years; edu = incomplete secondary schooling; race = Black African; prov = Western Cape.

in the rate at which the oldest adults (aged 60+) of both sexes both headed households and lived alone. Pension receipt is only coded positively if respondents are aged over 60, meaning this term should be interpreted as an interaction effect with the 60+ age category. Pension receipt increases the chance that people will head households, men more so than women. But female pensioners are significantly less likely to live alone than non-pensioners which is probably related to the important social security role grandmothers play in South Africa. Male pensioners were 2 percentage points more likely to be living alone, by contrast, by the End GHS period (this was statistically insignificant in the LPM but the logit model found the same-sized coefficient to be significant).

More highly-educated men are more likely to head households than the least-educated; and people with post-secondary education are most likely to live alone. However, well- and low-educated women are both more likely to head households compared to the base category of incomplete secondary education. Consistently across time periods, women with either primary or less or post-secondary education were both more likely to head households than those with incomplete secondary schooling. There is some strong change in the coefficients on race reflecting the increased propensity of the base group, Black Africans, to head households over the period. People living in informal dwellings also became increasingly likely to be heading their own households and living alone, for men in particular.

In sum, the model yields sensible reults based on our expectations from Chapter 5. Four key results are, firstly, the effect of marriage on household headship works in opposite directions for men and women, but the same direction for living alone. Secondly, the interaction of being not-married and employed is much more important for women than for men, showing marriage and livelihoods are more closely bundled up for women. Thirdly, the marital status effects were extremely strong for men, in particular. Fourthly, we noted that the marital status effects were strengthening over time, so that not-married people were more likely to head households by the End GHS period than they had been in the OHS period. Together, these four points suggest marital status already played a vital role in shaping household formation behaviour at the beginning of the period, and that this only strengthened by the end of the period.

6.6 Decomposition results: sources of change in household formation

The detailed decomposition output for the total period is reported in graph format in Figure 6.4 with headship in panel A and living alone in panel B. We only report the breakdown by Period 1 and 2 for headship, since it is the most interesting, in Figure 6.5. The detailed output is also reported in table format in Appendix B.3 along with results from the Yun (2004) decomposition on the logit models in Appendix B.4. Again, results are highly robust. Figure 6.4 reports the percentage point change in the headship rate that is attributable to the coefficient, endowment, or interaction effect of each variable. For example, the never-married coefficient effect was responsible for increasing the male headship rate by more than 5 percentage points whilst the endowment effect served to reduce headship by about 2.5 percentage points (a clearer interpretation will be made shortly). A black vertical reference line marks the overall change that is being decomposed (and reported in Table 6.1). This provides some sense of how important the effects are relative to the change. The effects of some variables are summed together to make the output easier to read. This has been done for the age and education categories (separately

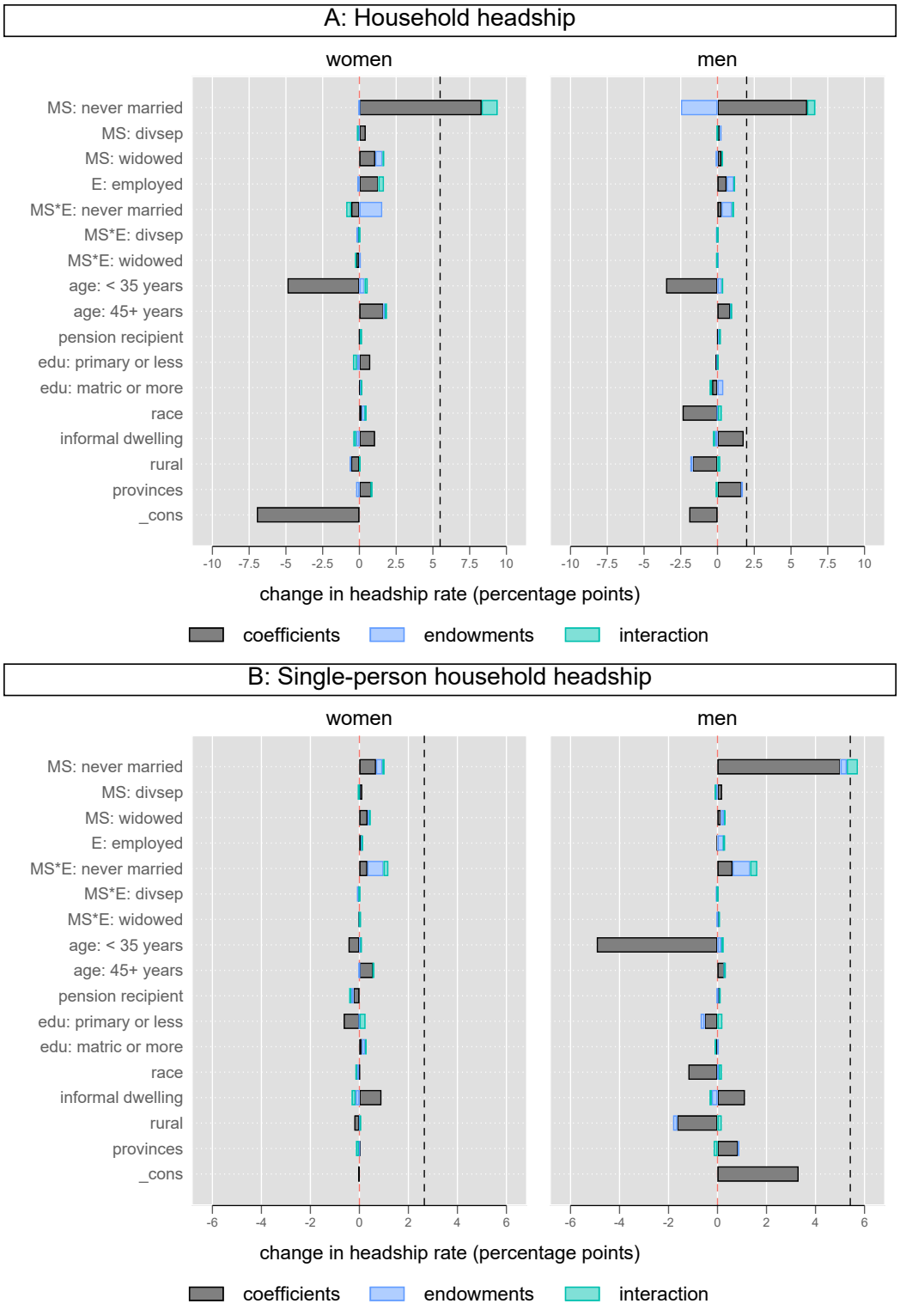
depending on which side of the base category they fall), as well as for race and provinces. Note also that the decomposition effects need to be interpreted in conjunction with how the underlying variable changed (i.e. did education levels go up or down over the period); the change in the predictors is reported in Table 6.5. The change in the coefficients is naturally the regression output discussed previously in Tables 6.3 and 6.4.

Parts of the findings from Chapter 5 are directly confirmed by the results in this section, whilst layers are added to our understanding of other parts. The part of the story from Chapter 5 that is strongly supported is that the marriage market has been an important driver of change. In Figures 6.4 and 6.5, the never-married effect is often the only one to ever account for more than the overall gap, by exceeding the black dotted line. What is also clear from the decomposition is that this is mainly happening via a behavioural channel because the large effects on never-married are coefficient effects, with hardly any contribution from an endowment effect. In Figure 6.4, the increase in the rate at which never-married women headed households would hypothetically have raised their headship rate by 8.32 percentage points (accounting for 152% of the gap), if composition was held constant. Never-married status also yielded an important coefficient effect for male heads and single-person households. These results strongly support a more behavioural interpretation of household proliferation. In terms of endowments, the 5 percentage point rise in the share of never-married men would have hypothetically *reduced* male headship since married men are much more likely to head households than never-married men, as the regression output just told us.

The part of the narrative from Chapter 5 that is developed further is the role of increased access to labour market income for women. Chapter 5 linked more female headship to more women gaining employment in the 1990s. In other words, we would expect this to play out in the form of a large endowment effect in Panel A of Figure 6.5, but this is not the case. Although there is a sizeable employment effect, this effect is dominated by a coefficient effect. This large coefficient effect on employment is one of the main factors differentiating women's headship boom in Period 1 from Period 2's plateau in Panel A of Figure 6.5. (The other differentiating factor is large coefficient and endowment effects on female widows in Period 1.) This suggests the boom was affected not only by more access, but importantly also by faster household formation by employed women. The regression coefficient on employment for women in the OHS period in Table 6.3 is actually slightly negative (confirmed by the logit in Appendix Table B.6). This suggests a fundamental shift in how women used their earnings to form households between the OHS and Start GHS periods. Acceleration in the rate at which employed women headed households over Period 1 would hypothetically have increased their headship rate by 1.77 percentage points, all else equal, accounting for 42% of the gap.

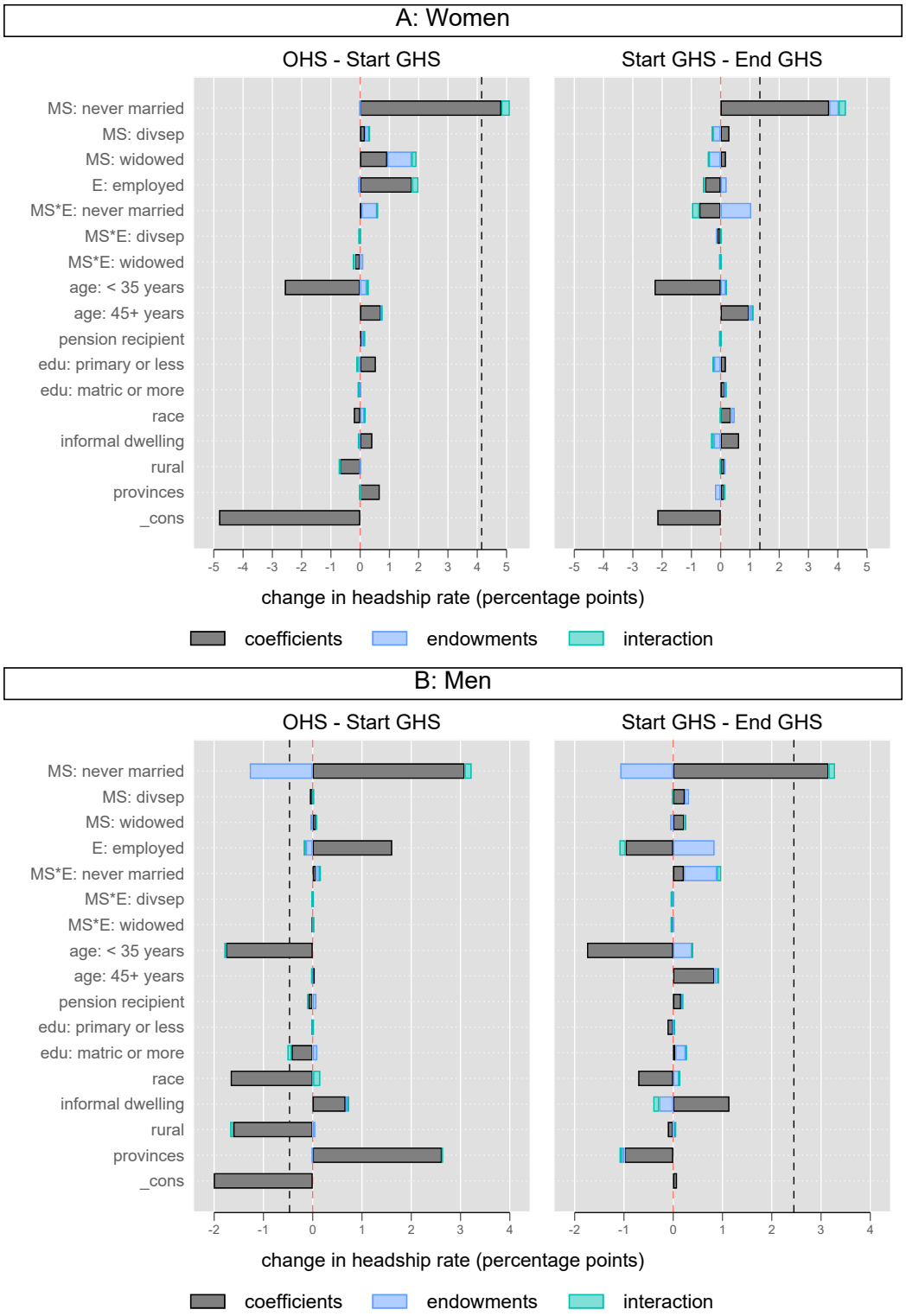
This behavioural story is different from an access story, which is instead probably operating via the interaction between employment and being never-married. The regression output informed us that most of the female employment effect was soaked up by its interaction with marital status. This makes sense because married women's access to labour market earnings is traditionally through their husbands. For women in Panel A of Figure 6.5, we now see a relatively large *endowment* effect on the never-married-employed interaction in both periods. In Period 2, the endowment effect on this interaction accounts for 78% of the gap, all else equal. The story about women's access to labour market income in other words is mainly playing out via the endowment effects on this interaction. At the same time, the decomposition has yielded the additional insight that there was also an important behavioural channel for employed

Figure 6.4: Detailed decomposition results for the Total period (OHS - End GHS)



Source: own calculations using the OHS-GHS series weighted using the Best CEW. Dashed black reference lines indicate the magnitude of the gap being decomposed as reported in Table 6.1.

Figure 6.5: Detailed decomposition results for the headship model in Period 1 (OHS - Start GHS) and Period 2 (Start GHS - End GHS)



Source: own calculations using the OHS-GHS series weighted using Best CEW. Dashed black reference lines indicate the magnitude of the gap being decomposed as reported in Table 6.1.

Table 6.5: Predictor means (% of population aged 15+ years)

	Women			Men		
	OHS 1996-8	**Start GHS 2002-4**	**End GHS 2009-11**	**OHS 1996-8**	**Start GHS 2002-4**	**End GHS 2009-11**
MS: never married	45.13	47.86	51.00	53.82	56.30	58.65
MS: divsep	3.66	4.03	3.40	2.31	2.28	2.02
MS: widowed	9.88	11.66	10.94	1.88	2.13	2.52
E: employed	26.25	29.53	33.74	44.02	42.80	48.36
MS*E: never married	9.24	11.16	14.88	14.21	14.90	20.82
MS*E: divsep	1.97	2.16	1.86	1.16	1.11	1.12
MS*E: widowed	1.85	2.48	2.41	0.49	0.59	0.79
age: 15-24 years	30.25	29.49	28.55	32.50	32.57	30.43
age: 25-34 years	23.74	22.99	23.56	25.37	25.38	26.72
age: 45-59 years	16.44	17.10	18.84	15.95	16.45	17.04
age: 60+ years	11.74	11.98	12.21	8.92	8.53	8.75
pension recipient	6.95	9.40	9.07	3.65	4.65	5.07
edu: primary school or less	39.49	34.03	24.81	37.30	32.59	23.25
edu: high school/matric	15.40	18.29	23.86	17.47	20.74	24.48
edu: post-secondary	6.21	7.42	9.47	7.28	7.28	9.41
race: coloured	9.31	9.13	9.24	9.21	9.13	9.02
race: asian/indian	2.77	2.72	2.75	2.93	2.91	3.02
race: white	12.98	11.88	10.72	13.83	12.28	10.97
informal dwelling	14.58	14.09	11.97	17.51	17.94	16.07
rural	42.04	42.24	36.69	37.69	39.19	33.72
prov: Eastern Cape	14.56	13.84	12.36	12.57	12.80	11.70
prov: Northern Cape	2.32	2.11	2.15	2.38	2.21	2.10
prov: Free State	6.50	5.94	5.53	7.06	6.28	5.30
prov: KwaZulu-Natal	21.76	21.23	20.16	19.90	19.42	19.02
prov: North West	6.53	6.77	6.35	7.07	7.35	7.03
prov: Gauteng	20.19	21.09	23.90	24.32	24.70	26.53
prov: Mpumalanga	7.16	7.16	7.54	7.29	7.21	7.46
prov: Limpopo	10.52	11.17	10.43	8.48	9.19	8.87

Notes: own calculations using the OHS-GHS series weighted using the Best CEW on the regression sample of respondents aged 15 and older. Base groups for categorical variables are as follows: MS = married; MS*E = married or not employed; age = 35-44 years; edu = incomplete secondary schooling; race = Black African; prov = Western Cape.

women in the boom period, which we might not otherwise have detected.

Being both never-married and employed was also an important combination for men's single-person households. Looking at men in Panel B of Figure 6.4, the interaction between never-married and employed yields a relatively large total effect of 1.65 percentage points (or, 29% of the gap). This total effect represents the next-largest effect contributing towards raising men's rate of living alone after never-married and the intercept term, to which we return shortly. In Panel B of Figure 6.5, men also have a large endowment effect on this interaction in Period 2, when men's headship rates took off. In other words, the growth in the share of people who are both never-married and employed (5.6 percentage points for women between the OHS and End GHS periods, and 6.6 percentage points for men) has contributed towards accelerating household formation.

The effects on pension receipt are practically zero. This is partly a function of the model and decomposition not being set up in a way to target the detection of the pension effect which is complicated by the age-eligibility cut-off. We have built a general model of household formation for the adult population, at large. We cannot definitively conclude from these results that the effect of the pension is insignificant

on the household formation of older people. Rather, within the ambit of a model for the adult population in general, we can say that there are more important drivers.

The change in the constant term often rivals the size of the largest individual effects. This change is the effect of group membership, which in our case is the change over time in the chance that our base category (married 35-44-year-olds) would head households or live alone. The chance a woman in this base group would head a household falls quite substantially over time by just under 7.5 percentage points; whereas the chance the base category man will live alone increases. It is impossible to interpret this change in a way that is neutral to the base category problem. This interpretation problem also affects the quite large coefficient effects on age. The slow-down in the household formation behaviour of young adults (relative to the base group of middle-aged adults) reduced household formation by up to 5 percentage points for women and 3 percentage points for men. Faster household formation by older people relative to middle aged, increased the headship rate by about a percent. Men and women ultimately had similar drivers of change, mainly being never-married, age, and the interaction between employment and never-married. Spatial variables (mainly informal dwelling) were sometimes more important for men and widowhood more important for women.

6.7 Conclusion

This chapter took a more formal approach to understanding the trends plotted in Chapter 5 and the benefit was to place Chapter 5's conclusions on a firmer foundation and gain some new insights in the process. One of the most important findings in this chapter, consistent across all groups and outcomes and robust to various specifications, was the acceleration in the rate at which never-married people formed households. Behaviour change, rather than the changing marital status composition of the population, emerges as the stronger effect in the decomposition. This conclusion is not immediately obvious from the trends that show a substantial increase in the share of never-married adults over time.

Why might this effect be so strong? Earlier in this chapter we described a possible mechanism, already supported by some South African evidence. In their study of the United States, Craigie et al. (2018) associated higher rates of household headship by never-married women with poorer local marital prospects, proxied by higher shares of unemployed men. If this mechanism is at play in South Africa, then very high rates of male unemployment could be one factor encouraging faster household formation by never-married women. In this instance, Posel and Casale (2013) have already made the link between high levels of Black African male unemployment in a local area and the lower chance that Black African women would be married. The present chapter does not go so far as to make the links that Craigie et al. (2018) have made. However, our work so far clearly shows that the marriage market is playing a vital role in changing household formation patterns in South Africa.

The trend analysis in Chapter 5 also associated the feminisation of the labour market with the boom-period of female headship in the 1990s. An insight from this chapter is to associate this change more closely with never-married women, emphasising how closely marriage and access to earnings are bound up together for women, in a way they are not necessarily for men. This effect cropped up both in the regression results and the decomposition. In the former this happened via large coefficients on the interaction between marital and labour market status that diluted the coefficients on female employment.

In the decomposition, the impact of the expansion of female employment in the 90s appeared to work almost entirely via the endowment effect on never-married employed women, specifically.

The decomposition captured the effect of the feminisation of the labour market in another unexpected way, which was a large coefficient effect on female employment, i.e. a behavioural change. During the boom period, there was a significant acceleration in the rate at which women leveraged their employment to form their own households. Expansion in women's freedom at the time, regarding reproductive rights and domestic abuse, for example, is one plausible reason why employed women might have changed how they spent their paychecks. Other than this coefficient effect on employment, the decomposition differentiated the boom period in female headship from the plateau period by a large total effect from widows. There was a notable increase in the share of women who were widowed, which is possibly related to the country increasingly coming under the grip of the HIV/AIDS epidemic. Both of these factors stabilised in the plateau period.

Here at the end of Part II, we have made some progress in developing an understanding of household formation in post-apartheid South Africa. Aside from the effects of population growth and aging, there has been notable change in household formation behaviour that warrants close attention from researchers. This change has been highly gendered and clearly related to how the labour and marriage market modulate men and women's means to support themselves, which in turn has vitally interacted with high rates of unemployment and falling rates of marriage. The result is that, perhaps in response to the optimism immediately after apartheid and along with a number of laws expanding women's freedom, women embarked on an expansion in household formation in the 90s. Women who were not married, but nevertheless had access to earnings, were a key group here. However, as the decade of the 2000s wore on, the reality that the labour market would remain exclusive set in, putting an upper limit of women's independence. Female household formation settled down at a rate matching adult population growth.

For men, the rise of the single-person household seems intimately related to the decline in marriage. At first this might have been only younger men, but as these men aged and still did not form a union, its plausible they continued to live alone. These gendered patterns are linked to men's greater ability to live alone. As the past two chapters have shown, employment (without marriage) is a more viable route to household formation for men than for women, because men continue to earn more in the labour market. Next, we turn to extending our understanding of how household formation change has taken place over the life-cycle in Part III, by studying the process of leaving the parental home. This is a natural focal point for household formation in general; but also makes sense given that marriage is a key event in the life-cycle.

Part III

Home-leaving

Chapter 7

Leaving the parental home in post-apartheid South Africa

7.1 Introduction

When noticing that household formation has increased, a common question people have is whether this is happening because young adults are leaving the parental home sooner. Strong intuitions about the household formation process over the life-cycle makes home-leaving a natural focal point, borne out in the economic literature on household formation which often focuses on young adults (Rosenzweig and Wolpin, 1993, 1994; Ermisch, 1981, 1999). For our study, household formation and home-leaving are related, but distinct concepts. For us, household formation is associated with the establishment of a new household linked to a household head. Young adults leaving home do not necessarily have to contribute to household formation if they join an already existing household, e.g. moving into their spouse's existing household, or move in with relatives. Nevertheless, faster home-leaving should contribute to more household formation (because it is unlikely all home-leavers will be joining existing households). On the other hand, if home-leaving is stable or even delayed and household formation has increased nevertheless, we should turn our attention to other factors to explain this increase. In South Africa, we know remarkably little about home-leaving beyond some original work by Simkins (2017*a*,*b*). This matters because our expectations about home-leaving trends are ambiguous at the outset based on what we know about the harsh economic conditions faced by youth, strongly motivating an empirical investigation into the topic.

When we introduced the thesis, one of our primary motivations was to understand how the household context had changed for South Africans over the post-apartheid period because of the importance of these social connections for a broad sweep of social outcomes. Youth are uniquely important in this regard, firstly because of their positioning to shape future socio-economic and demographic outcomes - in the period 1995-2011, the 15-35-year-old age group made up approximately 40% of the population throughout. But also because young adulthood is a 'demographically dense' period of life (Rindfuss,

1991) in which many important life events cluster, such as the end of education, beginning of work, leaving the parental home, union formation, and onset of parenthood. The timing and sequencing of these events at earlier ages have profound implications for other important events over the life course because life trajectories are highly contingent (Settersten, 2003). For example, an early pregnancy or protracted hiatus between the end of education and the first job has the potential to set individuals on completely different life trajectories. For these reasons, home-leaving and its relation to other events commonly understood as part of the 'transition to adulthood' have been the subject of a large and rich literature focused mainly on the West and also East Asia (Aassve et al., 2002; Billari, 2004; Raymo and Ono, 2007; Suzuki, 2001).

In South Africa though, we need to be sensitive to how home-leaving may manifest differently in different family patterns. For Western cultures, leaving home is a rite of passage associated with the transition to adulthood and governed by the rules of the nuclear family pattern (Ziehl, 2002). For agnatic extended family patterns in South Africa, the tradition is for young people to remain in the fold of the extended parental household for a few years after marriage and child-bearing before moving out with their spouse, children, and potentially other kin (Ziehl, 2002; Siqwana-Ndulo, 1998). Labour migration interrupted this ideal, however, as work drew young people out of the rural areas and to the urban areas (Seekings, 2008). In both of these family patterns, marriage triggers movement to a new stage. In the nuclear case this would be movement from either the parental household or a single-person household to the couple household. In the extended family pattern, this could mean a wife moving in with her husband's family.

Over the course of the past two chapters, we have associated the drop in marriage rates over the post-apartheid period with more never-married people forming their own households and especially living alone. Under the nuclear paradigm, worsening marital prospects might mean young men move out and live alone and then take longer to consolidate into a couple household. But, women might remain in the parental home for longer before either joining a couple household or forming their own single-person household, leading to an ambiguous net effect on household formation. Under the extended family pattern, the result of less marriage may not be detectable if young adults remain in the parental household anyway. If on the other hand, young adults are less interested in adhering to the traditional family formation of the past, or are drawn out of rural areas by the need to search for work, we might see more young people leaving the parental home. All in all, our expectations about home-leaving patterns are ambiguous at the outset.

We know very little about the macro levels and trends of home-leaving in South Africa since this topic has not received much scholarly attention beyond some original work by Charles Simkins (2017*a*,*b*). Simkins (2017*a*,*b*) used the information in the GHS to identify whether young adults live with their parents or not, and we use his approach in this chapter. Simkins (2017*a*,*b*) shows that in 2015 more young adults aged 15-35 years remained in the parental home (63%) than had left (37%) and investigates the alignment of home-leaver status with employment status, marital status, orphanhood, and migration. Also relevant are other conditions contextualising the lives of youth. Youth unemployment is especially high in South Africa. In 2017, 64% of 15-24 year-olds and 43% of 25-34 year-olds were unemployed by the expanded definition, compared to an overall expanded unemployment rate of 36% (DPRU, 2018). Furthermore, in 2018, about 37% of 15-29 year-olds were not in any kind of employment, education or training (NEET) and this group is predominantly female, Black African, urban and in the 20-29 year-old

age bracket (de Lannoy and Mudiriza, 2019). Most NEET youth are unemployed (48%) or discouraged workseekers (24%), although over 60% of NEET women classified as 'inactive' say they are occupied with childcare and housework (de Lannoy and Mudiriza, 2019).

These high levels of youth unemployment and NEET status are not mitigated by welfare spending. The two most important social grants provided by the State in South Africa are targeted at either the aged or the very young: the Old Age Pension and the Child Support Grant. The State makes no provision for young adults who are unemployed.[1] These tough economic conditions are part of why anthropologist Alcinda Honwana (2012) characterised South African youth as in a state of 'waithood' - a state of suspension between childhood and adulthood. According to this concept, young adults are too old to be considered children, but the binding constraints of their socio-economic circumstances restricts their ability to attain the usual markers of adulthood such as, financial independence, leaving the parental home, union formation, and fuller civic participation (Honwana, 2012, 2014).

This chapter uses the OHS-GHS series to answer questions about about when young adults leave home in South Africa; how this has changed over time; what the main reasons are for leaving home; where home-leavers are living; and, how this intersects with the rise in household formation. This chapter extends Simkins (2017*a*,*b*)'s work in three key ways: firstly, by applying his definition to the OHS-GHS series we are able to plot trends, and we can do this with the advantage of our updated survey weights from Part I. Secondly, we investigate the types of households home-leavers occupy in more detail; and, thirdly, we analyse the intersection between home-leaving and household formation. We find that most young adults remain at home under conditions of high youth unemployment and a lack of social grants targeting this age group that might make moving out more financially feasible. Home-leaving in South Africa happens relatively late; getting married almost guarantees women will leave home, but getting a job seems about as important as marriage for men.

Although there are important South Africa-specific factors at play, this pattern also remarkably resembles the home-leaving of Southern Europeans, described by Billari (2004) as a 'latest-late' pattern in contrast to the 'earliest-early' pattern of Nordic countries which we describe shortly in the next section. South Africa conforms with many of the socio-economic explanations for latest-late home-leaving, such as welfare spending that values the Family over the Individual (expressed as more pension versus unemployment grant spending); high levels of religiosity; and poorer labour market conditions. Differential home-leaving trends by gender suggest that a gentle increase in male home-leaving has translated into more young men forming new households, especially single-person households. By contrast, female home-leaving has been mainly stable and it is unlikely young women are major contributors to more female household formation.

The next section discusses the home-leaving literature in the developed and developing world. Section 7.3 explains our classification of home-leavers in the OHS-GHS series. Thereafter, follows a profile of home-leaving in South Africa in Section 7.4, describing who leaves home, when they leave, how this aligns with the marriage and labour market, and where home-leavers are living. We discuss how South Africa fits into the international literature on home-leaving in Section 7.5; and then circle back to considering our original question about how home-leaving intersects with household formation, in Section 7.6. Section 7.7 concludes.

[1]The State only provides very limited support through the Unemployment Insurance Fund which is funded through employer contributions, only lasts a few months and which obviously makes prior (formal sector) employment a pre-requisite which is not necessarily the case for many young adults in South Africa.

7.2 Literature review

7.2.1 Leaving home in the developed world

Since about the 1970s, the age of leaving the parental home has been increasing slightly in the West (Billari et al., 2001; Molgat, 2002; Evans, 2013) and Japan (Suzuki, 2001), and most clearly in Italy and Spain. This is partly attributed to increasing economic vulnerability because of globalisation, making it more important to prolong education and to delay family formation due to the demands of studying and work (Evans, 2013). The change has also been linked to ideational change stemming from the Second Demographic Transition (SDT) which emphasises the Individual over the Family (Lesthaeghe, 2010). This ideational change supports delaying family formation in favour of, for example, first achieving career goals or investing more in higher education.

Behind this overall pattern there is persistent regional variation which has been the topic of much research interest (Aassve et al., 2002; Billari et al., 2001). Billari (2004) analyses adulthood in a sample of European countries for 18-34 year olds. 'Adulthood' is defined as covering the three events of home-leaving, union formation, and the onset of parenthood. Two patterns emerge at the extremes, which Billari (2004) calls a 'latest-late' pattern versus an 'earliest-early' pattern. The earliest-early pattern is associated with Nordic countries where age is a key factor in home-leaving: most young adults detach from the parental home at around 19-20 years. Union formation tends to come later in life and is less obviously connected to home-leaving. By contrast, the latest-late pattern characterises Southern European countries where young adults stay at home longer and detachment from the parental home is more diffused across age. The median age of leaving home for cohorts born in 1960-65 was about 26 or 27 for men and about 23 for women in Spain and Italy (Billari et al., 2001). Instead, it is union formation that is more decisive in young people leaving the parental home.

These differences are linked to institutional and economic differences between these countries, often finding expression in the welfare state. The earliest-early pattern is associated with States putting stronger emphasis on individual rights, as proxied through higher spending on unemployment benefits and participation in social clubs (i.e. social life outside the family is vibrant). The latest-late pattern was associated with higher religiosity and with States that place more emphasis on the Family as an institution and traditional family values (Reher, 1998). This is proxied by less State spending on unemployment grants and strong spending on pensions instead (i.e. a prioritisation of older generations versus youth based on the important role the elderly play in maintaining strong family ties). Economic factors also play a role: labour markets in Nordic countries are generally better-functioning than in Southern Europe, implying young adults are less likely to need to rely on parents as a private social security net.

7.2.2 Leaving home in the developing world

Home-leaving has not been as well-documented in the growing literature on the transition-to-adulthood for the developing world, which has focused more on the onset of parenthood, union formation, and sexual debut (Sironi et al., 2020; Grant and Furstenberg, 2007), including one paper on South Africa by Goldberg (2013).[2] Instead, home-leaving in the developing world has mainly been the topic of individual

[2]Goldberg (2013) studies the events of school-leaving, first job, and first child; but, not leaving the parental home.

country case studies. The main theme emerging from this literature is heterogeneity. For example, in Uruguay (Ciganda and Gagnon, 2010), China and South Korea (Yi et al., 1994), home-leaving has been delayed, but has increased over time in Malaysia (Johnson and DaVanzo, 1998). Entering the labour market triggers leaving the parental home in Mexico (Pérez Amador, 2006), but in Northern Ethiopia (Ezra, 2000) and urban China (Ting and Chiu, 2002), marriage remains key. In Uruguay, those of both high and low socio-economic status delay home-leaving which Ciganda and Gagnon (2010) interpret as reproducing intergenerational inequalities. The same outcome in Mexico - intergenerational income inequality - was differently interpreted as being the result of a steep socio-economic gradient of home-leaving in which young adults of a lower socio-economic status leave home sooner (de Oliveira and Salas, 2008). Essentially, structural differences in political regimes, labour markets, marriage markets, educational attainment and cultural norms play a vital role in shaping local home-leaving behaviour. This makes the set-up we provided in Chapter 2 for the South African case of primary importance.

7.3 Identifying home-leavers in survey data

The data we use is the same stacked series of cross-sections in the OHS-GHS series the Best CEW weight from Part I. One limitation of using cross-sectional data is that we cannot track returns to the parental home. Many analyses of home-leaving from the the rest of the world use panel or retrospective data and home-leaving is identified as when a young adult no longer lives with their parents, given some age restriction defining adulthood (Ermisch, 1999; Aassve et al., 2002; Billari, 2004). South Africa does have panel data that could be used for these purposes, but these data either only begin more than a decade after the onset of democracy or are not nationally-representative.[3] The OHS-GHS series is, in fact, the only nationally-representative data source with sufficient detail about household relationships to chart trends in home-leaving since the beginning of the post-apartheid period.

To classify young adults as having left home or not, we follow rules defined by Simkins (2017*a*). Simkins (2017*a*) uses questions about parental mortality and co-residence as well as the 'relationship to the household head' question to define home leavers and stayers. In the GHS, and with some harmonisation in the OHS period, there are nine categories of relationship to the household head:

1. Head
2. Husband/wife/partner of head
3. Son/daughter/stepchild/adopted child of head
4. Brother/sister/stepbrother/stepsister of head
5. Father/mother/stepfather/stepmother of head
6. Grandparent/great grandparent of head
7. Grandchild/great grandchild of head

[3]Panel data from the South African Population Research Infrastructure Network (SAPRIN) is confined to three local areas in the provinces of Limpopo, Kwa-Zulu Natal and Mpumalanga. On the other hand, the National Income Dynamics Study (NIDS) data is nationally-representative, but the first wave only commenced in 2008.

8. Other relative of head (e.g. in-laws or aunts and uncles)

9. Person not related to the head

The sample is restricted to people aged 15 to 35 and an individual is considered to be staying at home if they meet one of the following conditions, as described by Simkins (2017*a*):

1. "At least one of the young person's biological parents belongs to the household.

2. The young person is a son/daughter/stepchild/adopted child of the head or a grandchild/great grandchild of the head.

3. The young person belongs to relationship category 4 (brother or sister of head), provided that there is neither a father nor a mother known to be alive and living in a separate household.

4. The young person belongs to relationship category 8 (other relative of head), is no older than 24 and is at least 10 years younger than the head and there is neither a father nor a mother known to be alive and living in a separate household."

Simkins (2017*a*) explains that the most difficult relation of the head category to allocate is category 8, other relatives of the head, since this is a catch-all category and could be very diverse. It could include older or younger generations of the household head. With condition 4, Simkins (2017*a*) is trying to capture the case where younger nieces, nephews, or other relatives have moved in with a head who is old enough to be a parental figure. All other young people are considered home-leavers. We follow Simkins (2017*a*) rules in all years but in addition, we allocate people to the home-leaver category if they are heads, spouses of the head, or parents or grandparents of the head. Note that parental co-residence and parental mortality are separate questions in the questionnaires of the OHS and the GHS (and not inferred from the household roster constructed from the relationships to the household head). However, the OHS years 1997-1999 lacked questions about parental co-residence and 1999 additionally lacked questions about parental mortality. We still include 1997-9 in our sample, but keep this problem in mind. Table 7.1 shows a breakdown of how household members' relation to the household head category is distributed by stayers and leavers along with the unweighted sample size.

Table 7.1: Composition of home-stayers and home-leavers in the OHS-GHS 1995-2011

	% Staying	% Leaving	N
head	0	100	119 034
spouse	0	100	65 990
child	100	0	329 753
grandchild	100	0	72 990
parent	0	100	718
grandparent	0	100	1 294
brother or sister	34.5	65.6	36 554
other relative	16.7	83.3	54 742
other household member	3.0	97.0	9 361
Total			690 436

Notes: own calculations based on a sample of 15-35 year-olds in the OHS-GHS series.

7.4 Profiling home-leaving in South Africa

7.4.1 Who leaves home in South Africa?

Table 7.2 reports home-leaving rates by gender and race for the two years at the beginning and end of the period under study and reveals some extensive changes. Between the the mid-90s and early 2010's, Black African men increased their rate of leaving home by eight percentage points and white men decreased it by ten percentage points. As a result of this substantial shift, home-leaving for Black African and white men converged on almost the same rate of about 43-45% in 2010/11. Women overall, leave home sooner than men, in line with the pattern for most European countries (Chiuri and Del Boca, 2010). Although whites were most likely to leave home both at the beginning and end of the period, Black Africans were the only group that increased their rates of home-leaving over time. The trend, based on research mainly in developed countries, is that younger people are staying at home longer in response to poor economic conditions in job and housing markets, reinforced by ideational change supportive of delayed family formation associated with the SDT (Billari and Wilson, 2001; Evans, 2013; Molgat, 2002). What then, is driving Black Africans, and especially Black African men, from home at younger ages? Part of this explanation may relate to unique political changes happening in South Africa that impacted people differently based on historical racial categories, as well as the region's long history of migrant labour.

Table 7.2: Percentages of youth aged 15-35 who have left home by gender and race with sample size in parentheses: 1995-6 and 2010-11

	1995-1996			2010-2011		
	Male	**Female**	**Overall**	**Male**	**Female**	**Overall**
Black African	35%	40%	37%	43%	43%	43%
	(26 289)	(31 204)	(57 493)	(28 263)	(29 736)	(57 999)
Coloured	37%	40%	39%	31%	37%	34%
	(4 040)	(4 610)	(8 650)	(3 420)	(3 622)	(7 042)
Asian	38%	51%	44%	39%	46%	43%
	(1 313)	(1 369)	(2 682)	(762)	(707)	(1 469)
White	55%	61%	58%	45%	54%	49%
	(3 138)	(3 327)	(6 465)	(1 485)	(1 441)	(2 926)
Overall	**37%**	**43%**	**40%**	**42%**	**44%**	**43%**
	(34 780)	**(40 510)**	**(75 290)**	**(33 930)**	**(35 506)**	**(69 436)**

Notes: own calculations using the OHS-GHS series weighted using the Best CEW.

Figure 7.1 reports the same information in Table 7.2 with more detail about the intervening years and underscores why plotting trends is important. We plot the share of home-leavers per year by gender and for Black Africans and whites. The greyed out section labelled 'ID issues' reminds the reader that these survey years lacked certain variables needed to identify home-leavers and we can expect home-leaving to be overestimated in these years. The figure shows that in the OHS period (even in years without identification issues) the rate of home-leaving changed quickly year-to-year followed by more stability in the GHS. The trends for white people are less stable in general because the sample is smaller. It appears though, that white men have reduced their rate of home-leaving over time; while white women have followed a U-shape. The changes in the 90s are very likely related to the first democratic election

in 1994, signalling an end to the white legal monopoly on certain human rights and other economic advantages which could have prompted more cautious home-leaving behaviour by white young adults. Similar patterns emerged for white male headship, in general, in Part II.

Black African men increased their rate of home-leaving immediately after apartheid ended in the 1990s. If the trend in the ID issues years are set aside, the trend for Black African women looks mainly level. The dip between 2002-2004 for both Black African men and women is out of step with the rest of the trend in the GHS period and could be related to the change in Master Sample that happened in 2005 (DataFirst, 2015). The trends look slightly more credible if we look between the OHS and the start of the GHS in 2005. The increase on the part of Black African men is possibly related to the expansion of political and economic freedoms in this period following the election in 1994.

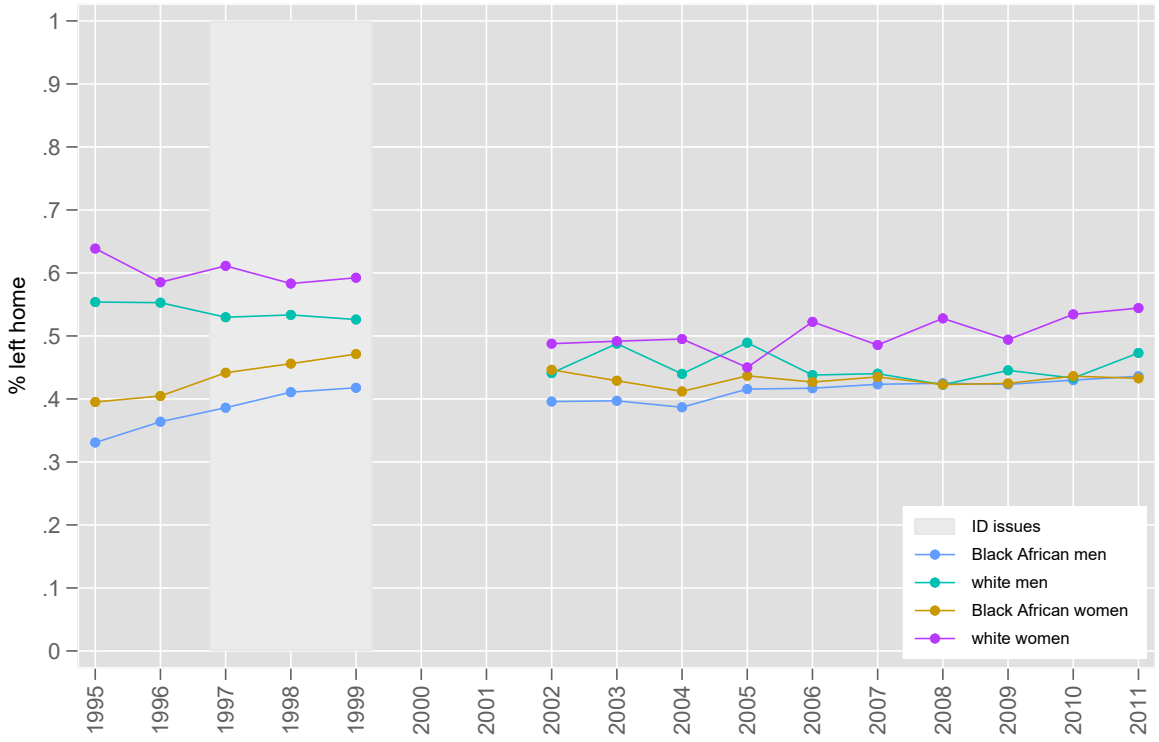
From 2005 onwards, home-leaving for Black African men and women converged. This convergence between male and female home-leaving for the Black African population is noteworthy and we are not aware of this close-to-complete convergence in other countries. Cross-country evidence from Chiuri and Del Boca (2010) reveals that the gender differential is smaller in Southern European countries than in Finland and Denmark, for example, suggesting the 'home-leaving gender gap' is another difference between the latest-late and earliest-early patterns. In their analysis of the gap, Chiuri and Del Boca (2010) conclude that women are more responsive to institutional factors, such as the labour and housing market. In South Africa, this could again relate to the provision (or lack of provision) of social grants for young adults. If Chiuri and Del Boca (2010)'s mechanism is correct, then the convergence in Figure 7.1 could be as much the consequence of slacker female home-leaving, as more positive male home-leaving.

Table 7.3 tries to establish how home-leavers look different to home-stayers and other adults in general. The table compares youth (15-35) who are home-leavers or stayers and older adults in the next 19-year age group (36-55). The table pools across years to portray general profiles. Home-leavers are unsurprisingly older than home-stayers and more home-stayers are full-time students. We investigate age in more detail in the next sub-section. Home-leavers, though, tend to be more educated than older adults which is in line with the expansion of access to education that happened over the post-apartheid period. This is mainly reflected in higher shares of high school graduates. NEET[4] status varies considerably by gender and home-leaver status. Male home-leavers are less-likely to be NEET whilst female home-leavers are more likely, resulting in an even bigger gender gap in NEET status amongst home-leavers: female home-leavers are NEET at about twice the rate of male home-leavers.

Although home-leavers are much more likely than home-stayers to be employed, female unemployment is relatively stable across home-leaver status. About a third of female youth are unemployed, regardless of whether they have remained or left the parental home. By contrast, 32% of men staying at home are unemployed compared to 21% who have left home. Young adults are more likely to be unemployed than older adults, about a fifth of whom are unemployed. Female home-leavers are much more likely to report being NEA (27%) than male-home-leavers (15%), but not any more likely to be full-time students (both at 14%), contributing to higher rates of NEET female home-leavers. There is large disparity in marital status across home-leaver status also influenced by point in the life-cycle. Home-stayers are almost universally never married; whereas about a quarter and a third of male and female home-leavers are married, respectively. Older adults are much more likely to be married than home-leavers, but

[4]NEET status is defined as individuals who are not enrolled full-time in an educational institution and not employed. In their profile of NEET youth in South Africa, de Lannoy and Mudiriza (2019) report that 36.5% of 15-29 year-olds were NEET in 2018; our figure for the same age group in 2011 is highly comparable at 36.7%.

Figure 7.1: The rate of leaving home over time for young adults in South Africa by race and gender, 1995-2011



Notes: own calculations using the OHS-GHS series weighted using Best CEW. Sample restricted to those aged 15-35. 'ID issues' = years where the survey data lacked information about parental mortality and co-residence needed to properly identify home-leavers.

home-leavers are more likely to be co-habiting than older adults.

Home-leavers are the group least reliant on grant income according to Table 7.3. Whereas about 27% of home-stayers and 14-15% of older adults co-reside with a pension recipient, only 4 and 7% of male and female home-leavers, respectively do the same. The CSG is more prolific for female home-leavers. About 22% of female home-leavers are living in households where a child is a CSG-recipient and this is comparable to 25% of older female adults but much less than 37% of female home-stayers. Instead, home-leaver households appear more reliant on labour market income. Home-leavers are less likely to be living in households with zero employed adults than home-stayers; and instead live in households with the highest ratios of employed to not-employed adults across all groups. For example, in male home-leaver households, there were 0.58 employed adults for every not-employed adult compared to 0.27 in the case of male home-stayer households. These data do not provide information about remittance income and the extent to which labour market and other income is flowing in and out of households and so our profile is limited in this way. Based only on what we can tell from Table 7.3, it looks like home-leavers are much more reliant on labour market income given their more limited access to grant income. Female home-leaver access to grant income comes predominantly via their children.

The importance of the CSG for female home-leavers is of course related to this group being more likely

to live with children than male home-leavers. Male home-leavers live with an average of 0.87 children compared to 1.61 in the homes of female home-leavers. These numbers are less than the number of children living with home-stayers and older adults, although in each case, women are living with more children than men. Relatedly, women live in bigger households across the spectrum; and home-leavers live in the smallest households. Whilst (by definition) no home-stayers are living alone, home-leavers are about double as likely to live alone than older adults: 28% of male-home leavers lived alone compared to 15% of older men; 13% of female home-leavers lived alone compared to 6% of older women. It is difficult to identify migrant labourers in these data as we have mentioned before, but the fact that home-leavers are more likely than both stayers and older adults to be living alone, living in urban locations, and living in informal dwellings speaks to typical characteristics of labour migrants.

7.4.2 When do South Africans leave home?

In this sub-section, we describe how home-leaving aligns with the three main drivers of household formation that have threaded throughout this thesis: getting older, getting a job, and getting married (Ermisch and Di Salvo, 1996; Ermisch, 1999). Figure 7.2 charts the age-specific rate of home-leaving for Black Africans and whites for two years at the beginning and end of the period of study. A key takeaway from this plot is that home-leaving is diffused across age for both Black Africans and whites (at least by the end period), so that South Africans could be described as 'late' home-leavers in general. This is particularly the case for Black Africans. Even by age 35, more than 20-25% of Black Africans still remained at home in line with extended family patterns of young people remaining in the extended parental household.

As many as 10% of Black African 15-year-olds had left home in both time periods. Greater home-leaving at this young age could point more to vulnerability rather than agency on the part of the home-leaver (Simkins, 2017*b*). South Africa was very much in the grip of the HIV/AIDS epidemic over this period and this result could represent orphanhood as a passage to 'home-leaving'. Temporary fostering is also common amongst Black Africans in order to send children to better schools or keep them out of unsafe ones (Russell, 2004), or because their parents need to migrate for work-related reasons (Hatch and Posel, 2018; Hall and Posel, 2019). The age-profiles for Black Africans shift left over time so that home-leavers start representing the majority after about age 25 in 2010-11; whereas this had been more like age 26.5 (women) or 28 (men) in the 90s. It was mainly younger men under the age of 30 who closed the gap between male and female home-leaving; so that by 2010-11, Black African men and women were leaving home at remarkably similar rates.

There is much more instability in the age-profiles for whites because of the smaller sample size. In the 90s, most whites had left home by about age 22; but, this shifted up to about age 23 for women and age 25 for men by 2010-11. Women maintained a higher rate of home-leaving than men over time and, if anything, widened the gap between the profiles from their early 20s onwards by 2010-11. The age-profiles for white home-leaving is much steeper than that for Black Africans in line with a nuclear family pattern emphasising the importance of establishing an independent household. By age 35, less than 10% of white young adults were still at home. Whites are also much more likely than Black Africans to be at home between ages 15-20, likely owed to their higher socio-economic status acting as a protective barrier to keep very young adults in the parental household during school-going years.

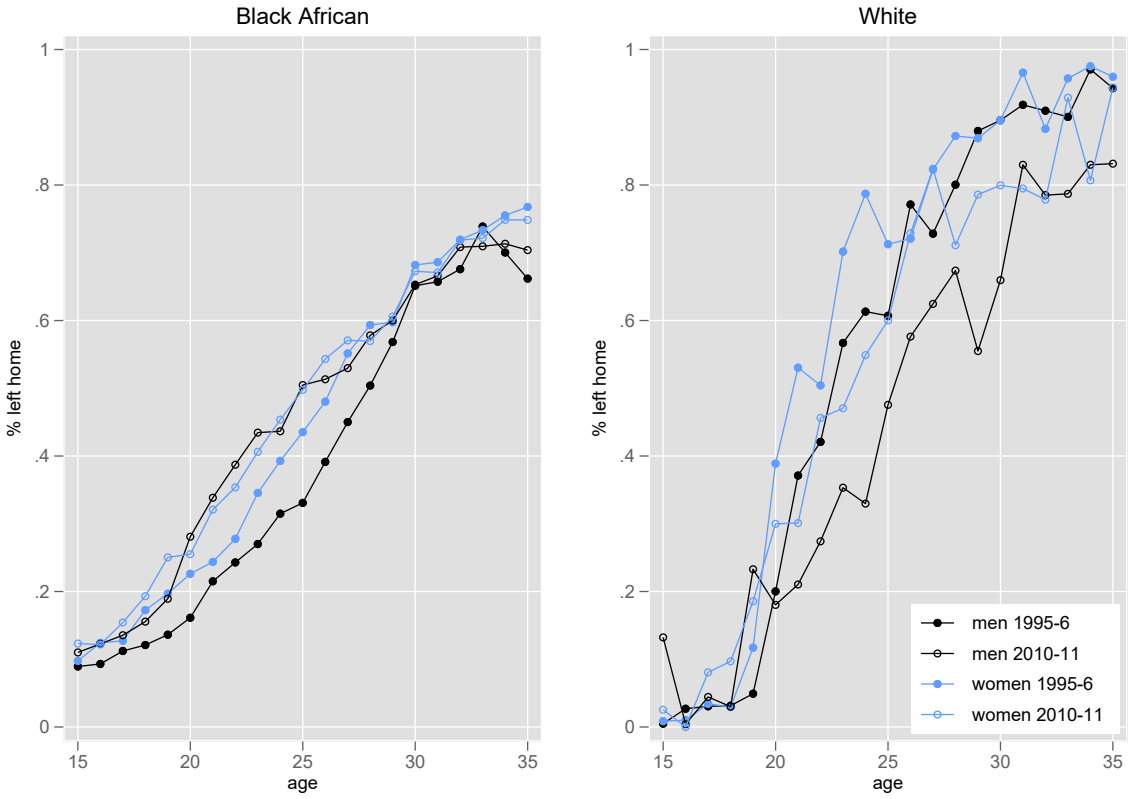
In Figure 7.3, we turn to age effects across generation. Recalling the overall trend in home-leaving

Table 7.3: Descriptive statistics of home-leavers, home-stayers, and older adults in South Africa, 1995-2011

	Stayers aged 15-35		Leavers aged 15-35		Adults aged 36-55	
	Men	**Women**	**Men**	**Women**	**Men**	**Women**
Individual Characteristics						
Age	22	22	27	27	44	44
Full-time student	45%	44%	14%	14%	1%	2%
NEET	36%	42%	23%	48%	31%	51%
Mental/Physical disablity	3%	2%	2%	2%	11%	9%
Education Level						
No schooling	2%	2%	3%	4%	10%	13%
Incomplete primary	13%	9%	11%	10%	21%	21%
Primary	8%	7%	6%	6%	8%	9%
Incomplete secondary	52%	55%	41%	43%	33%	32%
Secondary (matric)	22%	23%	29%	26%	18%	15%
Tertiary	4%	4%	10%	11%	11%	10%
	100%	100%	100%	100%	100%	100%
Labour Market Status						
Employed	20%	15%	65%	40%	69%	48%
Unemployed (broad)	32%	35%	21%	33%	19%	22%
NEA	49%	51%	15%	27%	12%	30%
	100%	100%	100%	100%	100%	100%
Marital Status						
Married	3%	2%	24%	35%	58%	52%
Cohabiting	1%	1%	13%	17%	11%	8%
Never Married	95%	96%	62%	44%	24%	23%
Divorced/separated	0%	1%	1%	2%	5%	7%
Widowed	0%	0%	0%	1%	2%	10%
	100%	100%	100%	100%	100%	100%
Household Characteristics						
Pensioner in HH	27%	27%	4%	7%	15%	14%
CSG receipient in HH	29%	37%	11%	22%	17%	25%
Other grant recipient in HH	14%	14%	4%	5%	10%	11%
No employed adults in HH	32%	34%	21%	28%	18%	26%
HH adult employment ratio	0.27	0.26	0.58	0.48	0.50	0.40
Household size	6.83	7.30	3.38	4.31	4.69	5.34
Single-person household	0%	0%	28%	13%	15%	6%
No. children (<15 yrs) in HH	1.97	2.47	0.87	1.61	1.37	1.74
Urban	56%	55%	71%	65%	69%	64%
Informal dwelling	10%	10%	33%	24%	19%	14%

Notes: own calculations using the OHS-GHS series weighted using the Best CEW. Sample pooled across time for all variables, except marital status. For marital status, GHS 2002-4 are excluded because these years did not differentiate between cohabiting as a separate category to married. HH adult employment ratio = ratio of employed to not-employed adults in the household.

Figure 7.2: Age-specific rates of home-leaving by gender and race over time for South Africa



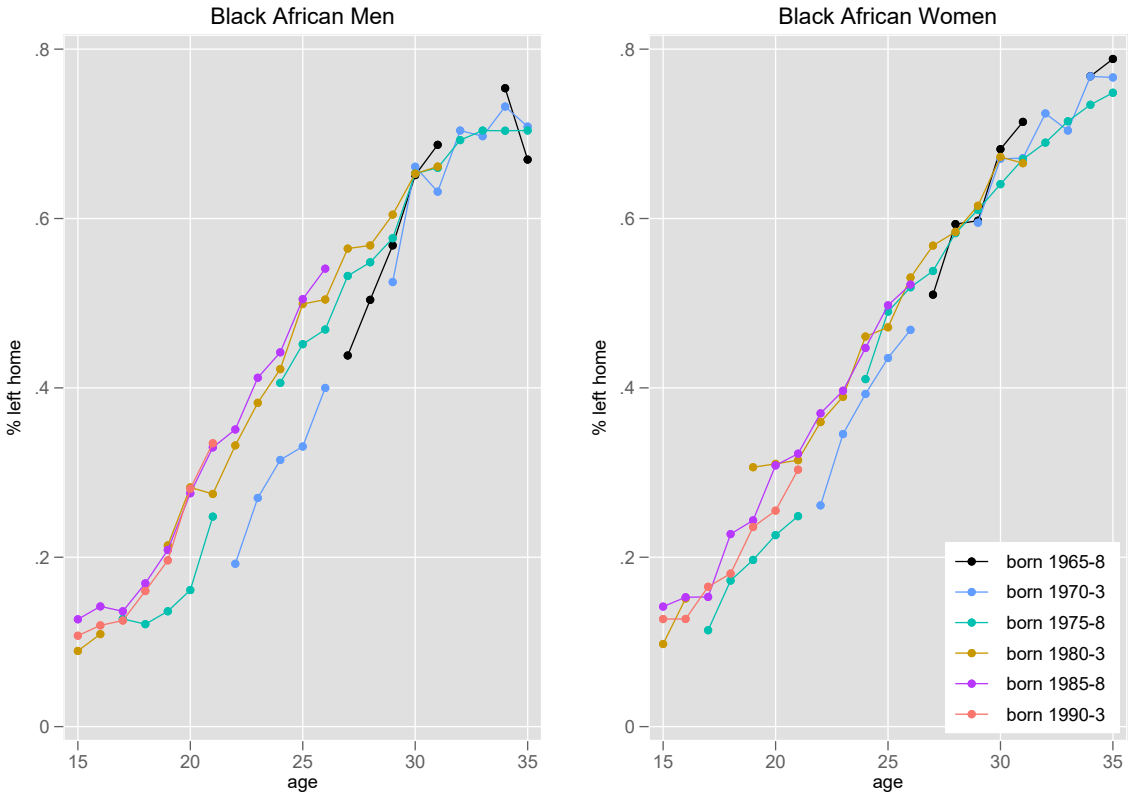
Notes: own calculations using the OHS-GHS series weighted using Best CEW. Sample restricted to those aged 15-35.

in Figure 7.1, one concern is that the structural break in the trend around 2000 and 2001 coincides with the change in survey instrument from the OHS to the GHS. The question is, to what extent is this structural break 'real' versus influenced by the change in survey? One way to separate out these effects is to plot age-specific rates of home-leaving by birth cohort, which has the added advantage of informing us about the extent to which home-leaving has evolved over generations. In Figure 7.3 we focus on Black Africans (because they are the majority of the population and their trend was distinct from other groups) and chart home-leaving across age for six three-year birth cohorts between 1965 and 1993. We pool across three years at a time when creating the cohorts to achieve more robust results given that there is already a gap in the data between years 2000 and 2001, and we additionally exclude years 1997-9 to avoid identification issues in these years biasing the results.

The results are largely assuring that the trends in Figure 7.1 are real. Immediately, we can observe the stability of home-leaving for women. Age-specific home-leaving rates for different cohorts lie practically on top of each other. There are larger changes for men in line with what Figure 7.1 reported. About 35% of Black African men born in 1970-3 had left home by age 25; but about 50% of those born in 1980-3 had left home by the same age. Cohorts born in 1975-8 and after are more similar to each other. The figure reveals that first it was older young adults (aged about 20-30) that started moving out sooner starting with the 1975-8 generation. But then younger young adults (aged about 18-23) ramped up their rate of

home-leaving even further, starting with about the 1980-3 generation and consolidating by the 1985-8 generation, the latter of which would have come of age about five years after apartheid ended.

Figure 7.3: Age-specific rates of home-leaving for Black African birth cohorts, 1995-2011



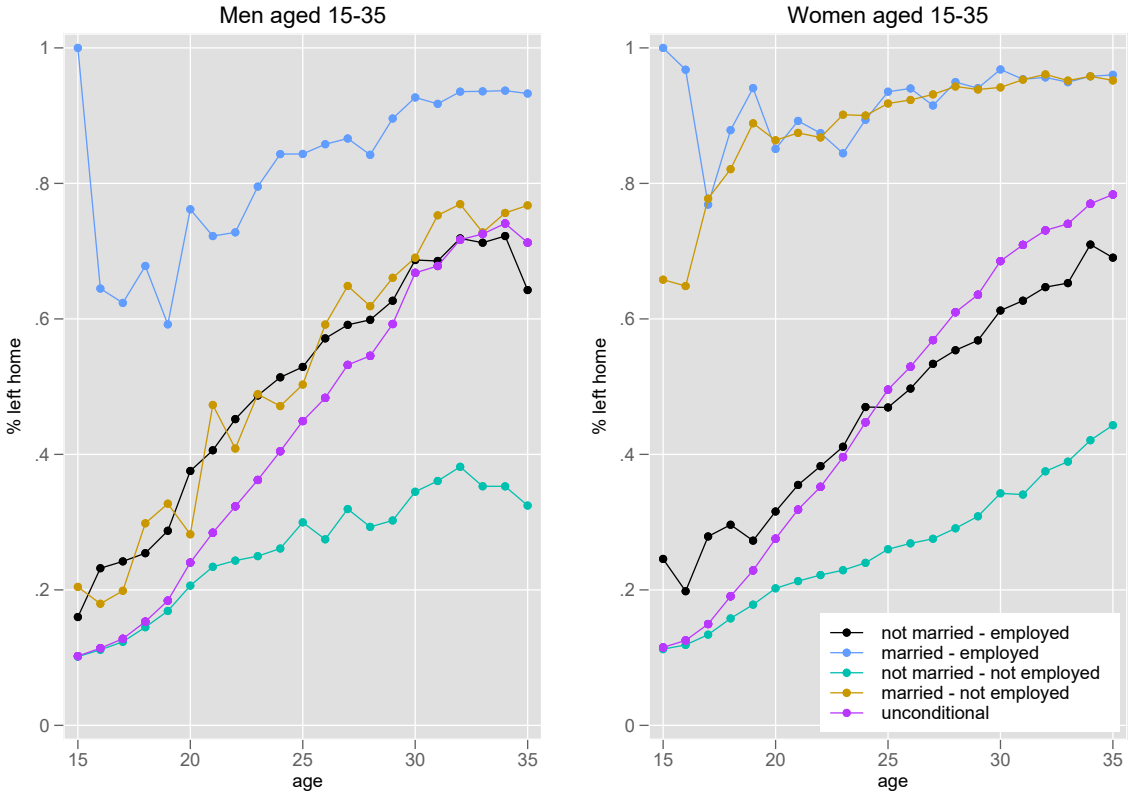
Notes: own calculations using the OHS-GHS series weighted using Best CEW. Sample restricted to those aged 15-35. Years on the legend denote year of birth. Years 1997-9 were excluded due risk they would bias the results because missing variables needed to identify home-leavers in these years mean home-leaving is overestimated.

Moving on to the role of marriage and employment, Figure 7.4 presents rates of home-leaving conditioned on marital and employment status. In this figure, 'not married' includes the statuses of never married, divorced or separated, and widowed. 'Married' then includes being either married or living together. 'Not employed' includes unemployed or not economically active (NEA) status. The share of home-leavers in combination categories of these variables are plotted against the unconditional rate of home-leaving over age. As we are plotting shares and not imposing a functional form, there is some instability in the estimates at young ages. This occurs simply because the samples get very small at these young ages, e.g. there are very few married and employed 15-year-old males. Naturally, this problem is eased as age increases.

Marriage is clearly predominant for women's home-leaving. Regardless of employment status, being married almost guarantees that women will leave home with home-leaving rates exceeding 80% almost always across age. Employment without marriage seems to spur some female home-leaving prior to age 25, but thereafter the unconditional rate of home-leaving exceeds this conditional rate. For men,

the interaction between marital and labour market status appears more important. The combination of marriage and employment stands out for making male home-leaving more likely than all other combination categories and the unconditional rate. Quite interestingly, the unconditional rate of male home-leaving then appears not that different to the rates for both 'married - not employed' men *and* 'not married - employed' men, although the former represents a small portion of home-leavers throughout. Either being married or being employed makes male home-leaving likely, but the combination makes it most likely by a substantial margin.

Figure 7.4: Age-specific rates of home-leaving in South Africa conditioned on labour market and marital status, 1995-2011



Notes: own calculations using the OHS-GHS series weighted using Best CEW. Sample restricted to those aged 15-35. Not married = never married, divorced/separated, or widowed. Married = married or cohabiting. Not employed = unemployed or not economically active. Employed = employed.

7.4.3 Where are home-leavers living and what are their household roles?

Figure 7.5 describes features of the destination households of home-leavers. In this figure, we describe two aspects. In panel A, we report the relationship of the home-leaver to the household head according to the categories in our survey data. The only amendment we make is to differentiate between married and cohabiting spouses, which necessitates the omission of data from 2002-2004. The questionnaire in these

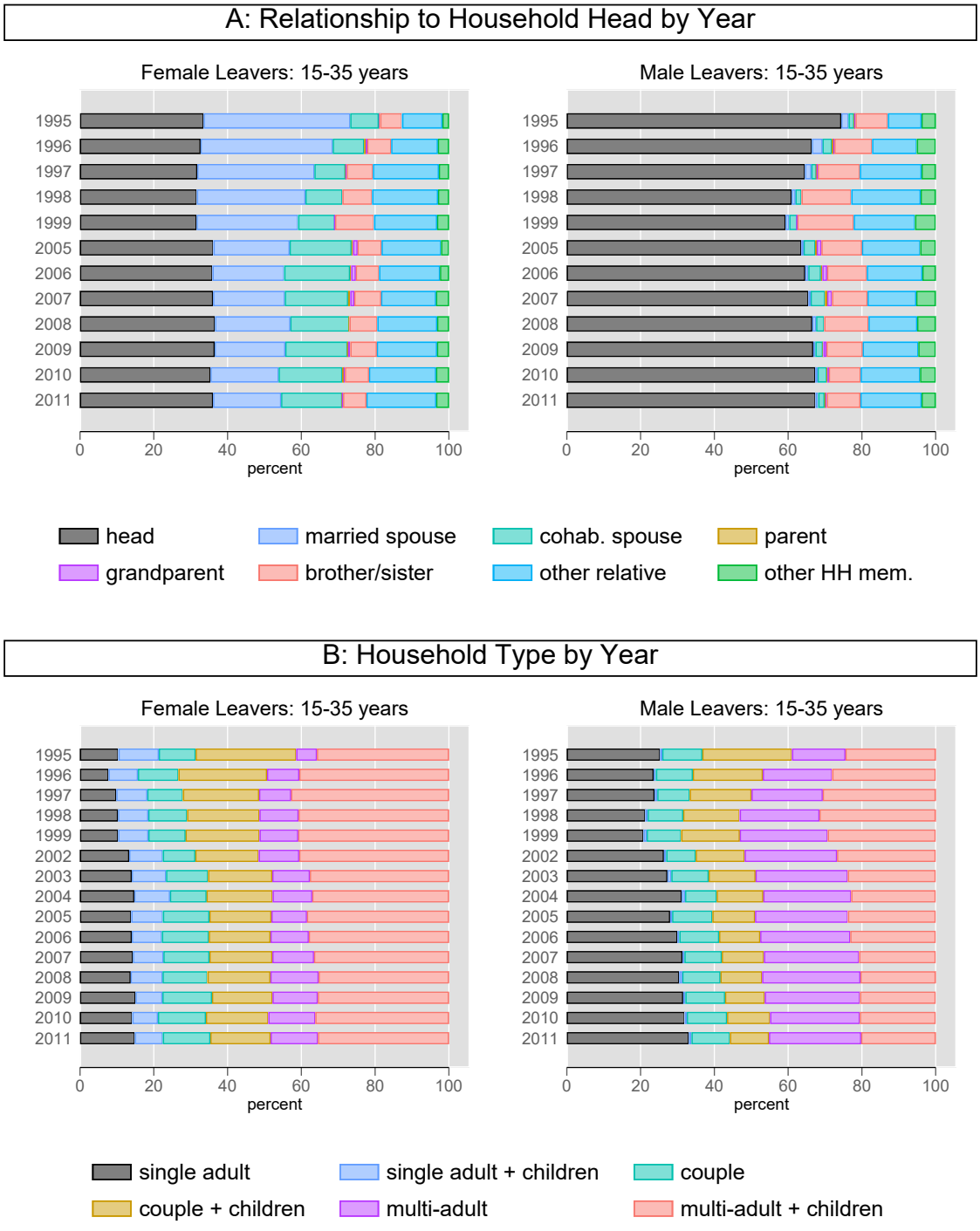
years did not separate out 'cohabitation' from 'married' when asking about marital status. In panel B, we classify households into six mutually exclusive household composition categories. The categorisation hinges on age to label combinations of adults and children (aged under 15 years) because we are interested in the life-cycle and consider adult working-age children as analytically different from young children primarily in need of nurturance. The categories are single adult; a single adult with children; a couple; a couple with children; a multi-adult household; and a multi-adult household with children.

Looking at panel A, the share of women leaving to be spouses declined overall, but an increasing share are cohabiting instead of getting married. Both the overall drop in spouses and more cohabitation is probably linked to declining rates of marriage over the period. Two other trends are likely consequences of this, as well: more women leaving to head their own households or live with other relatives. Between the mid-90s and mid-2000s, we observe a step-up in the share of women who leave to head their own households, whereafter this proportion remains stable at about 35%. About 11% of female home-leavers were co-residing with other relatives in 1995, but this had expanded to 19% by 2011; highlighting the continued importance of complex household structures in South Africa. This share also appears to bulge in 1997-9, but we remind the reader that these years suffer from identification issues for home-leaver status making the trend less reliable, especially for the 'other relative' allocation. Men are about twice as likely than women to leave to head their own households: 67% of male home-leavers were heads in 2011. The change for men in panel A is mainly comprised of first a drop and then an increase in the share of household heads. Again, some of the steep drop in headship is reported for problem years 1997-9. For male home-leavers, too, complex households remain relevant. In 1995, 9% of male home-leavers were living with other relatives but this share was 17% in 2011.

Panel B reports that the two household types that appear to have grown the most over the period for both male and female home-leavers are single-person households and multi-adult households without children. More single-living was especially important for men: by 2011, about 35% of male home-leavers were living alone. Although more female home-leavers also started living alone, the more extensive growth for this group came from women living in multi-adult households. This likely coheres with the increase in female home-leavers living with other relatives in panel A. The decline in any type of household with children aligns with the general decline in fertility, but there is a large gender imbalance with regard to who lives with children. Female home-leavers are much more likely than male to live with children. Hardly any men live alone with children, whereas this represents a relatively stable share of just under 10% of female home-leavers.

Figure 7.5 then contributes to the debate in the literature about whether South African household structure is becoming more nuclear. Growth is happening at the 'poles' of the household complexity spectrum for home-leavers: single-person but also multi-adult households have grown. In a context of high levels of migrant labour, it is hard to conclude that growth in single-person households is necessarily evidence of a trend towards more nuclear family patterns. Instead, both panels underscore the continued importance of complex household structures. Although the most classic nuclear household, the couple with children, has declined, this is mainly owed to declining fertility. The share of home-leavers living as couples has remained quite stable and the main compositional change has been the displacement of households of any type with children (except, lone care-giver) by both single-person and multi-adult households.

Figure 7.5: Characteristics of home-leaver destination households and their relation to the household head



Notes: own calculations using the OHS-GHS series weighted using Best CEW. Sample restricted to those aged 15-35. Children = aged less than 15 years.

7.5 South African home-leaving in international perspective

Having profiled home-leaving in South Africa, we now consider how the South African case compares internationally. In spite of regional and socio-economic differences, South Africa would appear to fit quite neatly into the 'latest-late' pattern of Southern Europe, both in terms of the focal points of age and union formation, as well as the institutional and cultural contexts (Billari, 2004). Rates of religiosity are high in South Africa (Chipkin and Leatt, 2011); there is no general unemployment benefit (aside from limited unemployment insurance); and the highest-paying social grant is the Old Age Pension. The Child Support Grant is the welfare support most likely to land in the hands of young adults and constitutes a very small amount, often not increasing with inflation and coming in at less than the official food poverty line in some years (van der Berg et al., 2010; StatsSA, 2008*b*). However, one reason we would expect home-leaving to be more delayed by comparison in South Africa anyway, is that it is home to extended family patterns. In this family pattern, departure from the parental typically happens at later ages or life stages (after marriage and child-bearing, for example) than in the conventional nuclear family pattern, anyway (Ziehl, 2002).

However, Billari and Wilson (2001) also discuss literature citing the period effect that young people are leaving home later and delaying parenthood in a rational response to weak economic conditions. Whilst this explains trends for non-Black African population groups in South Africa, it doesn't explain why young Black African men in particular are more likely to have left the parental home under the same economic conditions (even after the initial adjustment after the advent of democracy). Two additional points might explain this divergence from the Southern European case: childcare and migrant labour. The first point of departure is that owing to strong family values and welfare spending that supports this, childbearing in Italy, for example, almost entirely happens within the confines of marriage (Billari et al., 2001).

This is very much not the case in South Africa. Declining rates of marriage and a long history of migrant labour have intertwined to de-couple child-bearing from marriage (Hosegood et al., 2009; Madhavan et al., 2013). Although total fertility rates are low in South Africa, very young women have a relatively high fertility rate and a common pattern is for women to have a first child early in their lives out of wedlock, followed by another one much later on (Dorrington et al., 2004). Men are overwhelmingly less likely to be primary caregivers than women and 68.2% children under 15 years lived without co-resident father in 2014 (Hall and Posel, 2019). As such, male home-leaving in South Africa may be less reflective of men leaving home to start their own families.

Instead, South Africa has a long history of migrant labour which continues today albeit having evolved with changing circumstances over the post-apartheid period (Bank et al., 2020). Many young men and women are 'double-rooted', balancing urban livelihoods with maintaining family and cultural ties in rural areas. Despite efforts by the post-apartheid government to stimulate rural development, most economic opportunities remain in urban areas, attracting the young and able to the cities. The profile of male home-leaving presented so far suggests work could be a compelling reason why more young Black African men might be leaving home faster. Rising home-leaving on the part of Black African men could reflect labour migration, either as part of stretched households or simply the need to move to where the jobs are.

There are limits on the extent to which we can trace people across space and time in our data because it is cross-sectional. As one point of entry though, Table 7.4 reports the share of young adults who have

left home by rural and urban location. Young rural Black Africans are much more likely to be staying at home than their urban counterparts and this is likely related to the extended family pattern in which it is conventional for young people remain in the extended parental home. In 1995-6, 37% of young Black African women living in rural areas had left home, compared to 44% of the same group living in urban areas. Although there has been some moderate change in rural home-leaving patterns, the more extensive change has been in urban areas where both young men and women have increased the rate at which they leave home. Young Black African men living in urban areas were 10 percentage points more likely to have left home by 2010-11. Also important is that female trends work in opposite directions in different locations: rural women were *less* likely to have left home by 2010-11, but urban women, much more likely. These counteracting trends are probably behind the ultimately level trend in female home-leaving overall. Higher urban home-leaving by women is probably the result of urban women being both more likely to be employed and to be married in our data.

Table 7.4: Home-leaving by location and race, in 1995-6 and 2010-11

% 15-35 y.o. in ... area who have left home	**Rural**		**Urban**		**% 15-35 y.o. in urban areas in 2011**
	1995-6	**2010-11**	**1995-6**	**2010-11**	
Black African women	37.00	34.68	44.07	50.24	56.33
Black African men	27.77	31.11	42.68	52.24	57.71
White women	69.07	55.57	60.64	53.81	94.18
White men	57.17	41.73	55.16	45.56	92.62

Notes: Notes: own calculations using the OHS-GHS series weighted using Best CEW. Sample restricted to those aged 15-35.

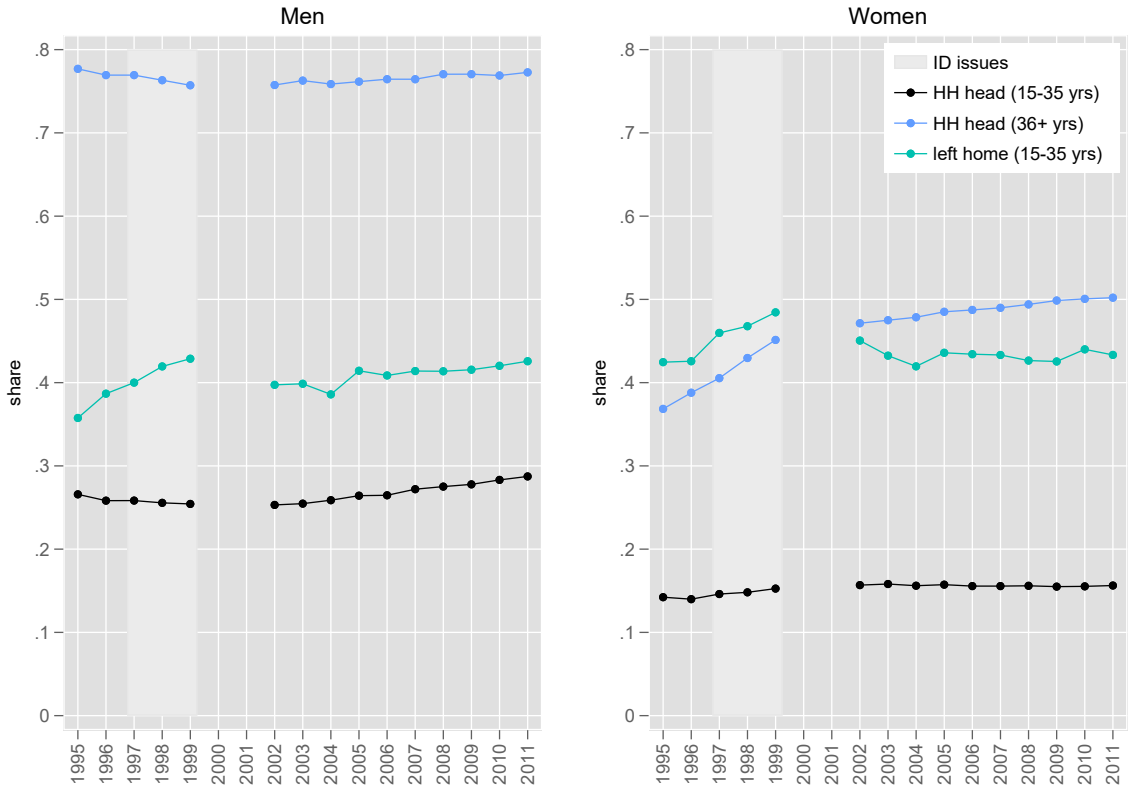
There are many possible explanations why urban Black Africans are leaving home sooner which we cannot separate out from these data: to escape crowded informal dwelling living arrangements; to better manage family relationships; migration out of rural areas to urban ones in search of work; stronger association with nuclear family formation ideals (Russell, 2003, 2004). On the other hand, the group most associated with the Western nuclear family pattern, whites, show different home-leaving patterns. Urban whites are the relevant group since the vast majority of white young adults are living in urban areas, and in a repeat of the patterns in the previous section, young white adults have become more likely to stay at home. This could well be a reaction to the end of protections for this group heralded by the end of apartheid. But, this delay is also more in line with global patterns of more young people delaying home-leaving in response to more precarious economic conditions, as well as, lengthening educational careers and delayed marriage (Billari and Wilson, 2001). Overall, there are lessons from the international literature on home-leaving that can help us understand the South African case. However, home-leaving in South Africa needs to be understood within the context of how the country's apartheid history impacted - and continues to impact - different groups.

7.6 How does home-leaving intersect with household formation?

We now turn to the question with which we opened this chapter, which is: to what degree might more home-leaving have contributed to faster household formation? One immediate reason we might think home-leavers are less likely to make an impact on household formation is that most young adults remain

at home. However, young adults make up a substantial portion of the population (40% throughout) meaning that even slight changes might have important cascading effects. Indeed, it turns out that the answer to this question varies by gender because different destinations and household roles belie similar aggregate rates of home-leaving for young men and women. Home-leaving patterns then are gendered, just like overall household formation patterns. To explore the intersection between household formation and home-leaving, we present Figure 7.6. We differentiate between the headship rates of young (15-35 yrs) and older (36+ yrs) adults in the black and blue lines, respectively. The green line is the trend in home-leaving, already presented by race in Figure 7.1. Because we have defined all heads as home-leavers, the black line is charting the share of young adults who have both left home and formed their own household.

Figure 7.6: Trends on the intersection of home-leaving and household headship



Notes: own calculations using the OHS-GHS series weighted using Best CEW. ID issues = years where the survey data lacked information about parental mortality and co-residence needed to properly identify home-leavers; HH head (15-35 yrs) = share of household heads amongst the 15-35 year-old age category; HH head (36+ yrs) = share of heads amongst the 36+ years age category; left home (15-35 yrs) = share of 15-35-year-olds who have left home.

The key result from this figure is that the headship rate for young men is inclining, whilst the same line is level for women. The headship rate for older men is inclining, but only marginally so, and less so than that for younger men. Older men are already highly likely to be heads, with their headship rate hovering just under 90% throughout the period. By contrast, there has been consistent expansion on the part of older women over the period, with the blue line clearly tracking the boom we described in Part

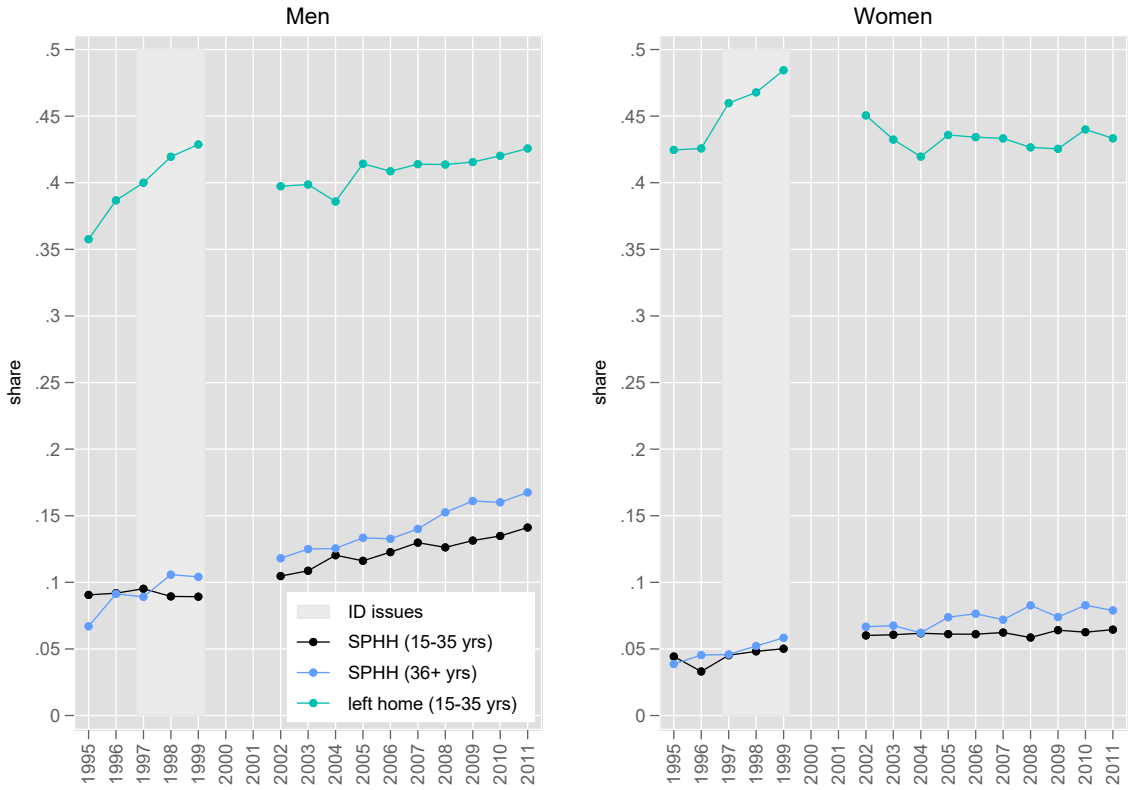
II in overall female headship in the 1990s, and then continuing to rise in the GHS period. The result is that the chance an older woman would head a household increased by 13 percentage points (or, 35%) in the 15 years between 1995 and 2011 so that by the end of the period, about half of older women were household heads. In other words, young men appear to be behind the growth in overall male household headship, but older women behind the growth in overall female headship.

These gendered trends appear to be the consequence of different trends in both the rates at which young men and women leave home; as well as, the rate at which these home-leavers then head their own households as opposed to joining existing households. Both the black line and the green line are increasing for men, and the black line is increasing slightly faster (a change of 3.4 and 2.8 percentage points respectively for the GHS period). This means young men are contributing to more household formation both because they are leaving home sooner, and because they are more likely to head their own households when they do. About 30% of young men (67% of home-leavers) had both left home and were heading their own households in 2011. Women on the other hand are more likely to join existing households than their their own households when they leave home, resulting in only 16% of young women in 2011 having both left home and formed their own household (slightly more than a third of home-leavers). There has been little change in either female home-leaving or the rate at which female home-leavers head households.

The same statistics are plotted for single-person household headship in Figure 7.7. In this case, both younger and older men appear to be contributing towards the growth of male single-person living, older men slightly more so than younger men. Again, the black line increases marginally more quickly than the green line (3.7 percentage points versus 2.8, respectively) meaning both men's faster detachment from the parental home and the faster pace at which they then split off into single-person households has contributed to more household formation amongst young men. Older women are behind the gentle increase in the rate at which women live alone in contrast to men and in a repeat of the pattern for women's headship in general.

Taken together, this suggests that faster male detachment from the parental home has contributed to household formation and living alone, but this is not the case for women for whom it is older women who are the vanguard. This coheres with findings from Part II that older women who were not married were increasingly likely to head households. This is a different profile to the female home-leaver who is more typically married: Table 7.3 reported that 52% of female home-leavers are either married or cohabiting with a partner. This emphasises a main theme about household formation, which is that the labour market and marriage market interact differently for men and women over the life-cycle.

Figure 7.7: Trends on the intersection of home-leaving and single-person household headship



Notes: own calculations using the OHS-GHS series weighted using Best CEW. ID issues = years where the survey data lacked information about parental mortality and co-residence needed to properly identify home-leavers; SPHH head (15-35 yrs) = share of people living alone amongst the 15-35 year-old age category; HH head (36+ yrs) = share of people living alone amongst the 36+ years age category; left home (15-35 yrs) = share of 15-35-year-olds who have left home.

7.7 Conclusion

The patterns of home-leaving in this chapter build on the findings emerging from the previous chapters by placing them more firmly within the life-cycle. Traditional family formation patterns in South Africa mean that historically Black Africans leave home later than whites since the former usually subscribe to an extended family pattern and the latter to a nuclear family pattern. This was reflected in the levels of home-leaving at the beginning of the period; but over time, Black African men started leaving home sooner, whilst whites started leaving later. The result was that both groups converged on about age 25 being when most young adults had left home in 2011. This still counts as relatively late compared to the developed world where young adults detach from the parental home as early as 18 or 19 years in some countries. Aside from family patterns, this late pattern can also be understood as a result of high youth unemployment and a lack of social grant support targeted at the youth.

A key part of the story that took shape in Part II was how change in household formation behaviour varied across age and gender. We found that prime-aged and older women were the group that increased their rate of household formation; whilst it was younger generations of men who appeared to be heading

more households and male solo-living rose across the age distribution. These findings make more sense now that this chapter has emphasised just how reliant young women are on marriage as a route out of the home. Since young women face such extreme unemployment rates, they are severely constrained in their ability to leave the parental home and set up their own household without the assistance of male earnings. Female home-leaving has been mainly stable over the period with more rural women remaining at home counterbalancing more women leaving home in urban areas - where jobs and marriage are more plentiful. It therefore makes sense that it is middle-aged and older women who have upped their household formation because, once women enter their 30s, they are much more likely to have a job. As such, faster household formation for women has mainly been driven by these older age groups and marital decline has likely played a key role in driving this increase as Part II has shown.

Men have more options to leave the home - they can get married, get a job, or both. This mirrors our findings for men and headship in general. Just as Black African men were the demographic who led the rise in household headship, they are also the only demographic to increase their rate of home-leaving over the period. As such, the gentle increase in Black African male home-leaving has contributed to more household formation. This is happening both because young men are detaching from the parental home sooner, and because they are more likely to either live alone or head their own household when they do. More work is needed to understand why young men would leave home sooner when the pattern around the world is for young people to delay home-leaving in response to weak economic conditions. The continuation of modern labour migration patterns could be one explanation behind this. It is well-established that people still travel vast distances for extended periods of time to work in South Africa (Bank et al., 2020). However, it is still not immediately obvious why this pattern would continue to increase over time even 17 years after the first democratic election.

Part IV

Implications and conclusions

Chapter 8

Some implications of changing household formation patterns

Parts II and III of this thesis have used seventeen-years-worth of survey data, carefully harmonised and reweighted in Part I, to provide insight into changing patterns of household formation in South Africa. In this chapter, we consider what the implications of these changes might be. This is done by either using our findings so far as a guide to explore some immediate touchpoints for household change in the OHS-GHS series; or, by reflecting and discussing what our findings might mean, more broadly. There are three sections in this chapter. The first section investigates how changing patterns of household formation have influenced trends in household composition. Within this, we specifically explore the composition of households headed by never-married people and how more single-person households might be influencing the pattern of household gender polarisation. In the second section, we consider what it might mean to have so many more people living alone. We discuss literature on some of the drivers and consequences of solo-living that locate the single-person household at the centre of critical social change of our time. Additionally, we consider what more single-person households might mean for the debate in the South African household literature about whether households are becoming more nuclear. Finally, the third section is a short discussion using the economic thinking set up at the beginning of Chapter 2 to structure our findings and explain how existing patterns of disadvantage are reproduced by household formation. As part of this process, we reflect on competing explanations in the literature for the overarching processes driving household change, with the goal of mainly raising questions to point out how central the household is for understanding modern social change.

8.1 Change in household composition

8.1.1 Growth at the poles of complexity

The key points emerging from the previous two Parts that we are most interested in capturing with respect to household composition are that household formation patterns are highly gendered and that there has been a mushrooming of single-person households. To plot change in household composition, we use the same classification we used in Chapter 7. This classification hinges on age as a defining feature for labelling household members as children (as opposed to relationship to the household head). This means adult children in a household with their parents are labelled as a household of adults. Whilst a different classification might label this a nuclear household, we think adult children who are of working age are analytically different to children primarily in need of nurturing from adult caregivers. Our conclusions are ultimately robust to classifying by relationship instead of age and our goals for now mean a simple measure of household complexity is sufficient. Figure 8.1 divides all households into one of six categories roughly increasing in complexity which we plot by gender of the household head: single adult; single adult with children aged less than 15 years; couples (i.e. household heads living with spouses); couples with children; multi-adult; and multi-adult with children.

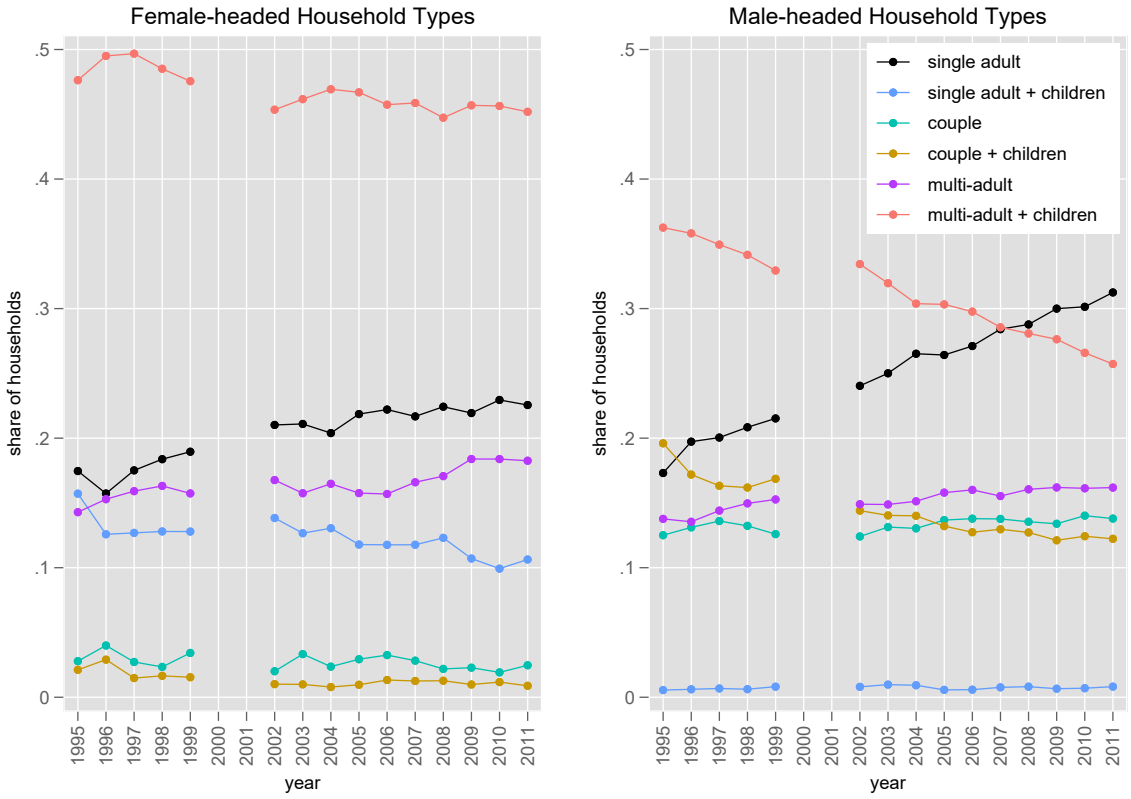
The profiles in Figure 8.1 diverge sharply by gender of the head. By 2011, the most common form of male-headed household was a single-person household; whereas this was a multi-adult household with children for female heads by a long margin. All household types with children declined as fertility rates have dropped, and multi-adult households with children declined especially quickly for men and were mainly displaced by single-person households. Single-person households have risen to quite comfortably be the second most-common household type for female heads and multi-adult households are the only other category to gain ground over time. Roughly equal shares of men and women head multi-adult households (which can include children of the head over age 15), but women remain likely to head up households alone with children while this is practically never the case for men. Overall, other than single-person households, the only other household type to increase is multi-adult, moderately for women and more gently for men.[1]

To investigate what these households look like in more detail, Table 8.1 gives more information about who makes up multi-person male- and female-headed households. We specifically limit the sample to multi-person households to separate out the growing influence of single-person households. For comparison and completeness, the same table is reproduced including single-person households in Appendix Table C.1. The surveys classify household members relative to their relationship to the household head using the following categories: head; spouse/partner of the head; child of the head; sibling of the head; parent of the head; grandchild of the head; other relative (e.g. in-laws, aunts); unrelated household member.[2] Table 8.1 then reports the share of households including at least one of each of these relationship

[1]This conclusion holds even if we assign children to households based only on their relationship to the head and not their age. In such a classification, adult children in the household would count as "children". The effect is to change the levels of some categories, but not the conclusion about the direction of trends: making this change results in a slightly flatter 'multi-adult' household trend and an increasing 'multi-adult + children' trend (couples remain level and 'couples + children' still decline). In other words, if you classify adult children as "children" the main change in terms of the trend is to transfer them from 'multi-adult' to 'multi-adult + children'. That is, complex or extended households are still the ones to increase. An interesting change in levels under this approach though is much higher levels of female-headed households living as lone parents to adult children.

[2]The categories include all instances where relations could be adoptive or step-relations. The 'grandchild' and 'grand-

Figure 8.1: Household compositional change by gender of the household head, 1995-2011



Notes: own calculations using the OHS-GHS series weighted using the Best CEW. Children = aged less than 15 years.

types, omitting heads themselves because 100% of households have heads. The only change we made to the survey question's classification was to differentiate between children and grandchildren of the head who were above or below age 15. These household membership categories are roughly classified by how they would fit into Goldscheider (1997)'s two axes of nuclear family formation - the male-female and the parent-child axis - with the addition of the 'Extended family' heading. Under the male-female axis we add a row for the share of households that include an adult member of the opposite sex to relate to insights about gender polarisation in South Africa (Posel and Hall, forthcoming).

The first row of Table 8.1 reports that almost all multi-person male-headed households included members of the opposite sex in 1995, and although this dropped by 6 percentage points by 2011, 88% of male-headed households still included women. In contrast, about 40% of multi-person female-headed households only include female adults and this did not change much over the time period.[3] There was, however, a clear drop in the share of households of either type that include a spouse.[4] Together, this suggests ambiguity in the direction of the male-female axis for South African households. This picture is an initial hint that the distribution of South Africans across households by gender is mediated in an

parent' categories include great grandchildren and -parents.

[3]Appendix Table C.1 on the other hand does register a small decline of 2.3%.

[4]We have not separated out same-sex couples because they make up such a small proportion of our sample. In 1995, a weighted share of 1.10% of couples were same-sex and by 2011 this had risen to 1.38%.

Table 8.1: Types of household members living in households by gender of the household head, 1995 and 2011: multi-person households

	Male-headed			Female-Headed		
% of households with at least one ...	**1995**	**2011**	**Diff**	**1995**	**2011**	**Diff**
Male-female axis						
Adult of the sex opposite to the head	94.3	88.1	**-6.1**	62.1	63.4	**1.3**
Spouse	88.5	79.0	**-9.5**	11.7	8.6	**-3.1**
Parent-child axis						
Adult child of the head (aged 15+)	46.2	41.5	**-4.7**	58.7	59.4	**0.7**
Child of the head (aged <15)	55.6	39.6	**-16.0**	51.4	35.4	**-16.0**
Parent of the head	5.5	1.8	**-3.7**	3.6	1.6	**-2.0**
Extended family						
Adult grandchild of the head (aged 15+)	3.9	5.8	**2.0**	12.1	18.2	**6.1**
Grandchild of the head (aged <15)	16.0	18.2	**2.1**	32.5	40.7	**8.2**
Grandparent of the head	1.2	0.1	**-1.0**	1.4	0.3	**-1.1**
Brother/Sister of the head	7.8	10.6	**2.8**	12.4	13.3	**0.9**
Other relative of the head	10.0	18.3	**8.3**	15.6	24.7	**9.0**
Other household member	3.6	3.1	**-0.4**	2.4	3.3	**1.0**
Total households (millions)	**4.80**	**5.81**	**1.0**	**2.71**	**4.60**	**1.9**

Notes: own calculations using the OHS-GHS series weighted using the Best CEW. Sample consists of all households of a size greater than one person.

important way by the growth of single-person households. We revisit this point in more detail in Section 8.1.3.

The fact that spouses have become less common has not been compensated for by a strengthening of the nuclear parent-child axis, which reduced across the board.[5] Instead, the dimension that has been gaining momentum is that of the extended family. Table 8.1 shows that the incidence of households including siblings and other relatives of the head increased by about 10 percentage points. Male-headed households went from 10% of households including other relatives in 1995 to 18.3% in 2011; the same change for female-headed households was 15.6% to 24.7%. In additional calculations, we found that most multi-person households include nuclear members[6] (92% in 1995 down to 85% in 2011), but the share of multi-person households including extended members (40% in 1995 up to 53% in 2011) is so large it can only be described as the norm in South Africa. The figures when single-person households are included are 76% of households including nuclear members in 1995 down to 61% in 2011; and 33% of households including extended members in 1995 up to 39% in 2011. In other words, the share of households with extended household members increased even in the face of rapid single-person household growth.

Higher proportions of female- versus male-headed multi-person households include siblings and other relatives of the head and are more likely to include adult children of the head and grandchildren of any age. (Higher proportions of both types of households including grandchildren could be a compositional outcome of reduced fertility.) Almost a quarter of female-headed multi-person households included other

[5]This is a crude lens on the nuclear family because these households could be three-generational, for example, which would not be a classic nuclear family. However, Posel and Hall (forthcoming) find that the incidence of three-generational households has been decreasing over time.

[6]'Nuclear' is defined according to Table 8.1, which could be leading to an overestimation because adult children might not strictly be considered nuclear household members, and households including both parents and children of the head at the same time would also not be nuclear.

relatives in 2011; and about 60% included either an adult or younger grandchild compared to a quarter of male-headed. Female-headed households are also more likely to include adult children than male-headed (approximately 60 versus 40% respectively), but it is unlikely that these adult children are living in an otherwise nuclear household set-up. Whilst most adult children living in male-headed households in 2011 had married parents (88%); the opposite was true for adult children living in female-headed households. The majority were living with widowed (43.75%) and never-married (27.31%) heads and a smaller proportion living with divorced or separated heads (10.13%).

The table and figure combined show that composition has grown at the poles of complexity and in a highly gendered way: by 2011, men were mainly forming single-person households, whilst women were mainly forming complex households that include children and extended kin. To the extent that this is undermining women and children's access to male earnings and time, this pattern is likely reinforcing women's double burden of work and childcare. Although these female-headed households could also include men living with children, we know from results elsewhere in this thesis (Table 7.3, Chapter 7) and other studies that men are less likely to co-reside with children (Hall and Posel, 2019); and even when they do, they are still less likely to take on primary care responsibilities (Hatch and Posel, 2018). So far, we have not reflected on whether these changes mean South African households are becoming more nuclear, specifically. This is because our conclusion in this regard depends almost entirely on how single-person households are classified, making it matter what we mean by 'nuclear'. We elaborate on these issues in Section 8.2.2. The points we make now, however, can be made without getting into this complexity: more people are living alone *and* more people are living with extended kin.

8.1.2 What types of households are never-married people forming?

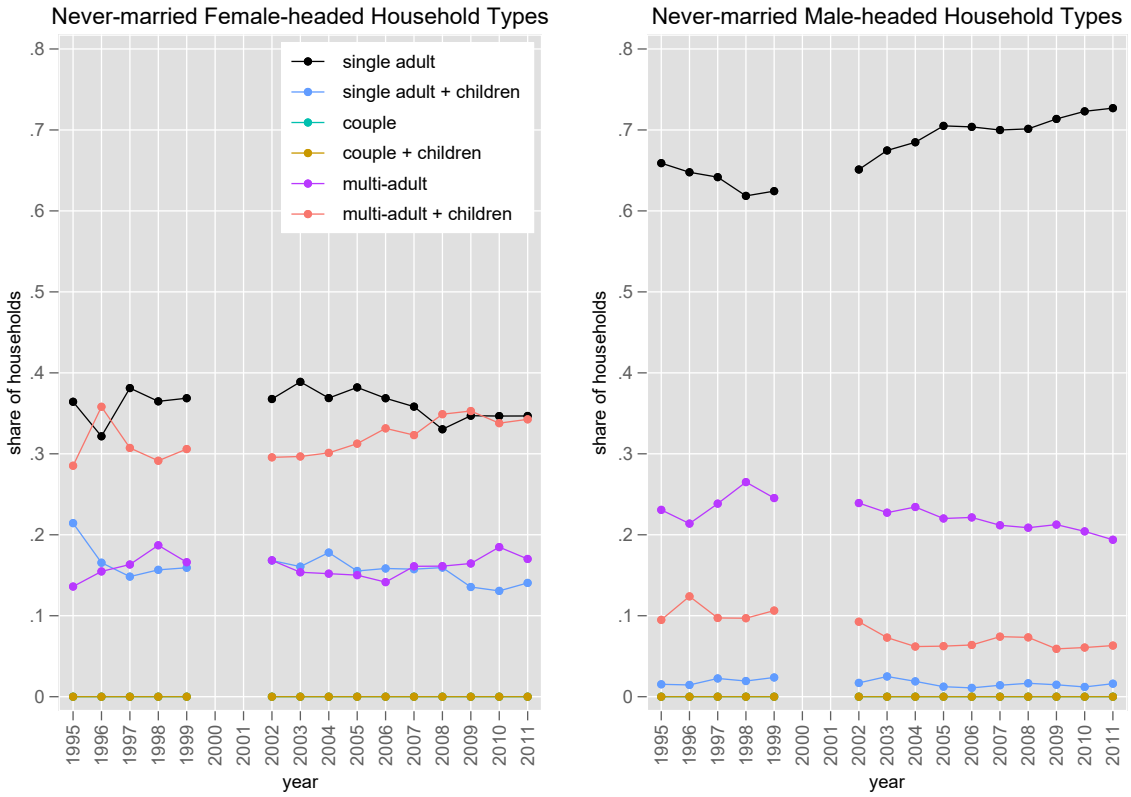
The rise of the never-married head was a central finding emerging over the course of the previous chapters. This makes it particularly interesting to investigate the types of households never-married people are forming, who lives in these households, and how this might have changed over time. Figure 8.2 and Table 8.2 in this section reproduce the previous figure and table (Figure 8.1 and Table 8.1) from above specifically for never-married heads. Again, Table 8.2 only uses multi-person households as the sample and a version with all households is in Appendix Table C.2. Although we often associated never-married heads with male single-person households in particular, this is not the only type of household never-married people form. Between 1995 and 2011, the share of multi-person households headed by never-married people doubled from 12 to 23%. By gender, the change was 6 to 12% of male-headed and 22 to 36% of female-headed multi-person households (own calculations).

What is clear from Figure 8.2 is that never-married men and women are forming quite different types of households to heads in general. Never-married men are mainly heading single-person households; by 2008, more than 70% of households headed by never-married men were single-person.[7] In notable contrast to Figure 8.1, the share of households headed by never-married female heads that include children (multi-adult with children) is actually increasing. This is quite remarkable given that this trend is strong enough to overpower the downward pull of falling fertility. This speaks to how women are adjusting coping mechanisms to meet new challenges of reduced access to male earnings but children's continued

[7]The result is that 90% of households headed by never-married men are occupied by male-adults only in an extreme version of household gender polarisation reported in Appendix Table C.2

need for care.

Figure 8.2: Household compositional change by gender of never-married household heads, 1995-2011



Notes: own calculations using the OHS-GHS series weighted using the Best CEW. Children = aged less than 15 years.

Turning to multi-person households in Table 8.2, households headed by never-married people are unsurprisingly more likely to only include adults of one sex and extended kin compared to multi-person households overall. Roughly 55% of both never-married male- and female-headed multi-person households were single-sex in 1995. More strongly than the overall case, Table 8.2 also shows that gender polarisation moves in opposite directions for never-married male- and female-headed households over time. When living with other people, never-married male heads are more likely to only live with other men, often brothers and other relatives. By contrast, never-married women heads are more likely to live with members of the opposite sex over time,[8] potentially a result of them increasingly including more adult children and grandchildren in their households. Households headed by never-married men are several times less likely to include children of any kind compared to those headed by never-married women, and compared to male heads of other marital statuses in the pooled sample in Table 8.1. Households headed by never-married women differ to the pooled sample in Table 8.1 by being more likely to include young children of the head - which may be behind the increasing share of multi-adult households with children in Figure 8.2. Both male and female never-married heads are more likely than heads in general to be

[8]This is true even when single-person households are included in the sample in Appendix Table C.2

living with siblings, other relatives, and other non-related household members. Siblings appear especially important for men, but both men and women saw growth amongst extended household members so that by 2011, siblings and other relatives were equally prevalent in households headed by never-married women.

Table 8.2: Types of household members living in households headed by never-married men and women, 1995 and 2011: multi-person households

	Never-married male-headed			**Never-married female-headed**		
% of households with at least one ...	**1995**	**2011**	**Diff**	**1995**	**2011**	**Diff**
Male-female axis						
Adult of the sex opposite to the head	44.1	36.5	**-7.5**	44.8	49.6	**4.8**
Spouse	0.0	0.0	**0.0**	0.0	0.0	**0.0**
Parent-child axis						
Adult child of the head (aged 15+)	7.4	9.2	**1.8**	35.1	44.7	**9.6**
Child of the head (aged <15)	10.0	6.1	**-3.9**	60.5	48.4	**-12.1**
Parent of the head	17.1	2.4	**-14.7**	7.3	1.5	**-5.8**
Extended family						
Adult grandchild of the head (aged 15+)	1.3	2.5	**1.2**	3.1	6.9	**3.8**
Grandchild of the head (aged <15)	2.9	3.8	**0.8**	14.5	24.7	**10.2**
Grandparent of the head	1.5	0.3	**-1.3**	1.5	0.3	**-1.2**
Brother/Sister of the head	53.9	55.0	**1.1**	29.9	28.1	**-1.7**
Other relative of the head	26.9	40.1	**13.2**	18.1	29.8	**11.8**
Other household member	18.2	11.7	**-6.5**	3.4	4.1	**0.7**
Total households (millions)	**0.26**	**0.70**	**0.4**	**0.59**	**1.67**	**1.1**

Notes: own calculations using the OHS-GHS series weighted using the Best CEW. Sample consists of all households of a size greater than one person.

As never-married heads grow as a group, the trends in this section could be considered especially relevant to shaping future household composition. If we were to adopt this idea, it means men are mainly forming households where they live alone, and if not they are forming households where they live with other relatives, often siblings. Women are forming households that typically include children of some type, but also extended family. These households are much more likely to be comprised of adults of only one sex. These patterns would serve to reinforce women's double burden of work and childcare. However, we must consider that the types of households never-married people are forming are especially in flux as the profile of both the average never-married person and never-married head changes as marital rates and fertility declines.

8.1.3 Household gender polarisation and living alone

Posel and Hall (forthcoming) identify growing household gender polarisation as an important characteristic of household change in post-apartheid South Africa. They report that in 1995, 26% of households were home to adults of one gender only and by 2018, this had risen to 46%. This trend matters because who one lives with is a strong predictor of the structure of one's livelihood strategies. Household compositional changes whereby female-headed households have become increasingly reliant on female versus

male labour market earnings (when women continue to be at a disadvantage in the labour market) is one reason why the poverty risk differential between male- and female-headed households widened over the post-apartheid period, to the disadvantage of the latter (Rogan, 2013). Household gender polarisation then has implications for evolving livelihood strategies and the extent to which new household forms are either challenging or reproducing existing forms of disadvantage, like women's double burden of work and childcare mentioned in the section above.

Our analysis of the reweighted OHS-GHS series finds similar levels of household gender polarisation to Posel and Hall (forthcoming); however, our special focus on single-person households, which are necessarily single-sex, reveals that the story is slightly more complicated. This insight is made possible by our careful reweighting work in Part I, enabling us to use the OHS-GHS series to investigate how single-person households shape this process in a reliable way in Figure 8.3. This figure differentiates between mixed-gender households (in monochrome) and single-sex households (in colour, green for women, blue for men); and also distinguishes between whether households are single-sex because they are single-person (SPHH) or because they are multi-person households (MPHH) in which all the adults are the same sex. The share of mixed-gender monotone households is steeply declining in Panel A when all households are included, but falls and then stabilises in Panel B when single-person households are excluded.

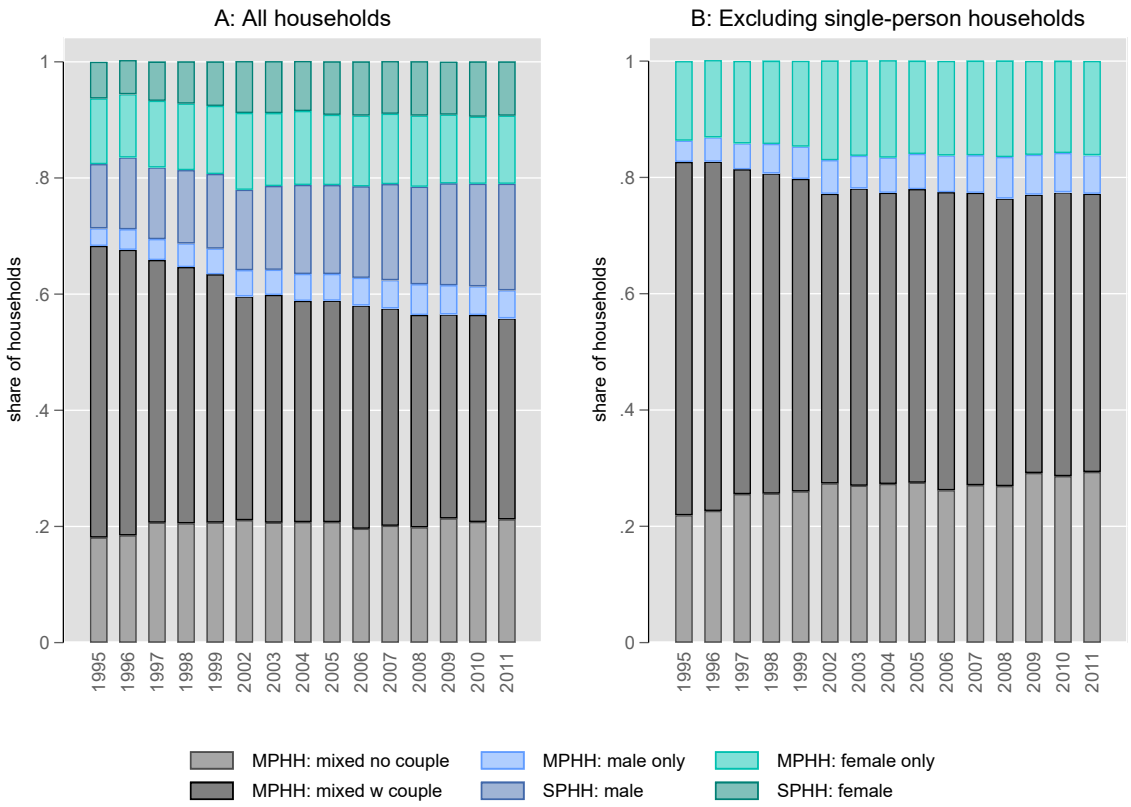
Once single-person households are excluded, the trend is that gender polarisation did increase in the OHS period, but then stabilised in the GHS period with about 77% of multi-person households being mixed. In other words, the household gender polarisation that we continue to see in this period in Panel A is entirely driven by the widening share of single-person households captured by the darker blue and green portions of the bars, and more specifically, men living alone. This matters for essentially the same reasons we would care about household gender polarisation in general, but changes the discussion slightly because living with other people of the same sex is different to living without anyone at all. As we make a case for in Section 8.2.1 later on, there are a slew of welfare considerations that are unique to living alone that make the growth of single-person households an important demographic phenomenon in its own right.

More single-living and more household gender polarisation are both likely influenced in a critical way by declining rates of marriage. Within the monotone mixed household share, we also differentiate between households that do and do not include a conjugal couple, based on insights from Posel and Hall (forthcoming). Looking at Panel A, the colour portion mainly eats into the share of mixed households with couples, whilst the share of mixed households without a couple only increases very slightly. In Panel B, living in a mixed household without a couple continued to increase while single-sex households largely stabilised and was as prevalent as living in a single-sex households.[9] What this suggests is that as marital rates have fallen, more people have started to live alone. But, when they are living with other people, living in a mixed-sex household with other adults is as common an alternative living arrangement as living in a single-sex household.

Household structure is key to understanding who does what in the household in terms of earning an income, caring for children, doing household chores, and usually has direct implications for livelihood strategies. The patterns of changing household formation then have implications for the degree to which household change is reproducing or challenging existing patterns of disadvantage. This section overall

[9]Male only and female only households together made up about 22% of households in 2011 in Panel B, while mixed households with no couple made up 29%.

Figure 8.3: Gender composition of multi- and single-person households in South Africa, 1995-2011



Notes: own calculations using the OHS-GHS series weighted using Best CEW. MPHH = multi-person household; SPHH = single-person household; mixed = male and female adults co-residing; couple = head and a spouse co-residing.

has shown that it is very likely that household formation patterns where men mainly form single-person households, but women form complex households including children and extended kin, serve to reinforce women's double burden of work and childcare. This is especially because this pattern is typified by the types of households the growing portion of never-married people form. The result is a consolidation of the bulk of childcare as women's work and the need to draw on extended family members, also often women, to assist with childcare and strategise about livelihoods. Men and women are also less likely over time to live in the same households as discussed in this section. To the extent that this means women are cut-off from male earnings, this adds to the pressure of the double-burden. In many cases, this pattern is driven by more people living alone. This means not only are men and women less likely to live with each other, they are less likely to live with anyone at all, which brings with it unique considerations, which we turn to next.

8.2 What is the meaning of so many more people living alone?

8.2.1 Drivers and implications of solo-living

Part of why we think it's important that single-person households are analytically separate from other household types is that their defining features make their drivers and implications unique in a number of different dimensions. In South Africa, one obvious reason for this is the association between migrant labourers and single-person households (Xulu, 2014; Smit, 2001). More generally, Jamieson et al. (2009) point out that women are often framed as drivers of social change because of the important interplay between marital change, women entering the labour market, and fertility reductions related to contraception in the Western world in the latter half of the 20th century. But it is men who have very much been the vanguard of the rapid rise of solo-living for a number of economic and social reasons applicable both in South Africa and internationally.

Firstly, men can afford to live alone more so than women; men continue to earn more than women in the South African labour market (Mosomi, 2019) and the cost of living alone is relatively expensive, a point to which we return shortly. Secondly, living alone can also be risky for women when inviting intimate partners into the home. This is true in a physical sense in South Africa's context of extreme levels of gender-based violence (Abrahams et al., 2009; Karim and Baxter, 2016), but thirdly also as Jamieson et al. (2009) point out in terms of reputational damage for women in a way that is not the same for men. A fourth factor is that women overwhelmingly take on the burden of childcare, making them much less likely to live alone (Hatch and Posel, 2018).

Figure 8.4 reports the education level of all adults aged 18 years and older[10] in 2011 categorised by which type of households they are living in. The figure shows that women who live alone differ to men who live alone both because they are more educated and because, compared to women living in other household types, they are more or equally as educated.[11] This possibly means that for women, some protection in the form of higher socio-economic status may play an important role in overcoming some of the barriers to living alone that women face.

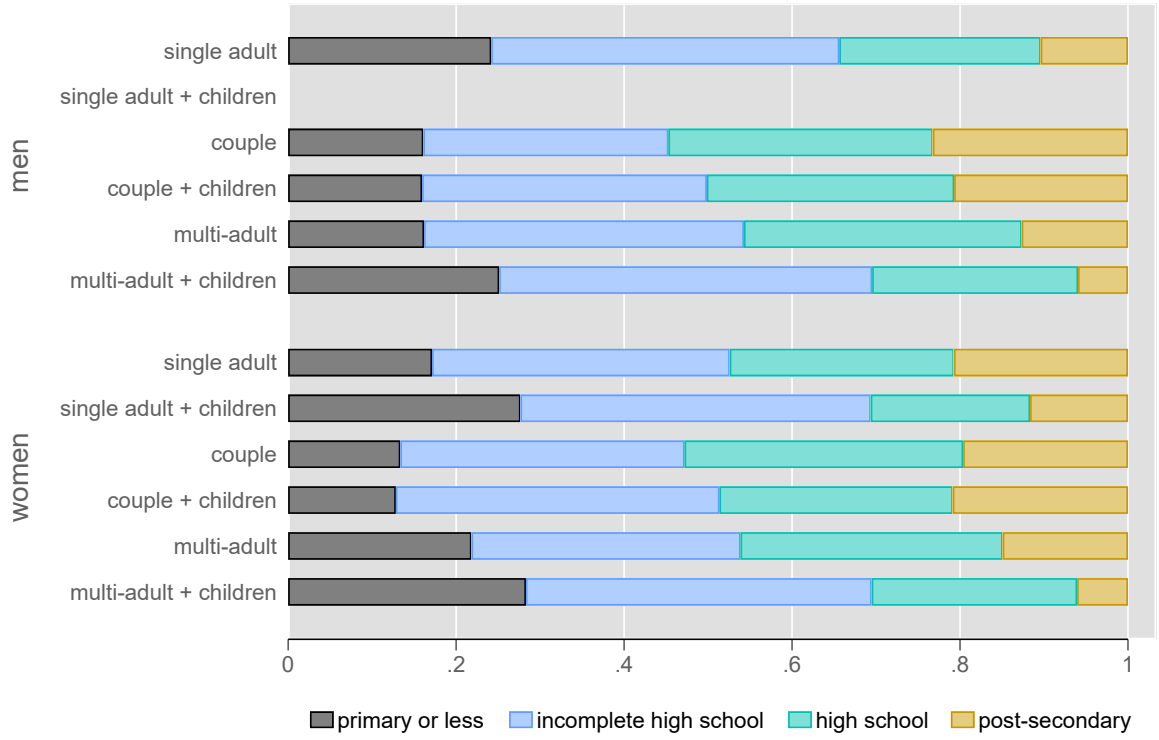
Living alone also has distinctive welfare outcomes. In health, for example, a medical literature is developing confirming the health risks posed by social isolation and loneliness. A meta-review by Pantell et al. (2013) concluded that social isolation is as important a mortality risk as traditional factors, like heart disease. Another meta-review of 148 studies by Holt-Lunstad et al. (2010) found that consistent across age, sex, initial health status, and cause of death, people with stronger social relationships had a 50% stronger likelihood of survival. Of course, social isolation, loneliness and living alone are distinct concepts which do not necessarily imply each other (Klinenberg, 2016). Some people who live alone are well-embedded in vibrant social networks, and some people who have been married for decades are nevertheless extremely lonely and socially isolated.

The review by Holt-Lunstad et al. (2010) noted their results varied by how social isolation was measured. Complex measures of social integration based on surveys asking questions about quality of social relationships and frequency of interaction, for example, had the strongest results which registered odds ratios in favour of socially integrated people of 1.91 (with a 95% confidence interval of 1.63-2.23).

[10] We changed our usual definition of adult from 15 years to 18 years because people usually complete high school at 18 in South Africa.

[11] This result holds if the sample is limited to household heads only.

Figure 8.4: Education levels for adults aged 18+ years by household type in 2011



Notes: own calculations using the OHS-GHS series weighted using Best CEW. Men living in 'single adult + children' households were omitted because they comprise such a small portion of the sample at less than half a percent of all male adults over 18 years.

By contrast, simpler measures yielded weaker effects. An indicator for not-living-alone increased the likelihood of survival by 19%; but the 95% confidence interval for the odds ratio ranged from 0.99 (i.e. no effect) to 1.44. Ultimately though, living alone correlates with scoring highly on the dimensions of social isolation or loneliness. A recent study of English adults aged 50 and older using factor analysis to try to separate out these conceptual strands found that both the 'moderately isolated and lonely' cluster and the 'highly isolated and lonely' cluster had high likelihood of living alone (Smith and Victor, 2019). Over 85% of the former and 50% of the latter lived alone. At the same time, a third cluster of people who lived alone scored highly for social interaction with children, other relatives, friends, and civic involvement.

Living alone also has implications for economic welfare. A well-established result in the literature on poverty is that people living in bigger households are more likely to be poor than those living in smaller households. This is often the case for poverty measures that classify people as poor or not by comparing their per capita household income to a chosen poverty line. People living alone often emerge as 'best-off' in this case because they do not have to share their income with anyone else. However, by focusing only on money coming into the household, this approach ignores the complexities that arise from spending the household budget. Larger households have better economies of scale, meaning that the cost of household public goods (e.g. housing) can be shared by more people so that living in a larger household could be more cost-effective than living alone (Fafchamps and Quisumbing, 2007). Using data from Pakistan,

Lanjouw and Ravallion (1995) show that the direction and strength of the empirical relationship between household size and poverty is highly sensitive to differences in the assumed size elasticity, choice of welfare measure, and the treatment of sub-groups like adults and children. The result is that the stylised fact that larger households are poorer could be an artefact of the assumptions that go into common methods of measuring poverty and, as Lanjouw and Ravallion (1995) point out, there is still no preferred method for making person-level comparisons across households of different size.[12]

In South Africa, Posel and Rogan (2016) use the 2008/9 South African Living Conditions Survey and find single-person households are least likely to be poor compared to bigger households when using household per capita income as the basis of poverty measurement. However, when using a question asking about own-assessment of economic well-being, multi-person households were often less likely to count as poor compared to one-person households. In other words, it is not immediately obvious that more single-person households mean an automatic improvement in economic welfare. Living alone will, in particular, be less cost-effective when the costs of household public goods are high, because these are costs that would be split across other members in a multi-person household (Lanjouw and Ravallion, 1995). In South Africa, there is evidence that costs of living are high with food price inflation in particular setting in after the Global Financial Crisis in 2008 (Leibbrandt et al., 2016; Mhlongo and Daniels, 2013).

The social and economic implications of having so many more people living alone are ambiguous at present. Studies from the developed world emphasise that more solo living is associated with higher levels of loneliness and anomie, but also more self-determination and freedom from potentially restrictive traditional social customs of the past (Jamieson et al., 2009). Clearly the single-person household is an outcome of a complex combination of socio-economic and demographic forces in the modern age, complicated by South Africa's own history. For this reason, we raise the question of what the single-person household might really mean for research on family patterns and household structure.

8.2.2 Are South African households becoming more nuclear?

As we discussed in the literature review in Chapter 2, during the first half of the post-apartheid period, scholars debated whether the trend of declining household size represented the gravitation of South African households towards the nuclear family form. Many researchers were more specifically interested in the question of whether Black African households, which have historically adhered to an extended consanguinal family pattern, were becoming more nuclear. Before we reflect on what our findings might mean for these questions, it is important to point out that researchers used the term 'nuclear' to mean slightly different things, and how this term was used could either change the conclusion or result in researchers talking past each other (Ziehl, 2002). There are broadly two approaches: the first approach is classifying *household structure* at a given point of time, something commonly associated with family sociologists and demographers. This approach usually focuses on whether the household is only comprised of nuclear family members - spouses, children, and the head. As soon as any other type of member is resident in the household, that household is considered an extended or complex household.

The second approach mainly taken by anthropologists is to classify *family patterns*; that is, to think about how the rules for family formation voiced by members of specific groups might result in different

[12]Using adult equivalence scales or incorporating a size elasticity parameter still requires researchers to commit to a picking a level for these variables about which researchers may disagree or which may vary by locality.

household structures over the life-cycle. Ziehl (2002) describes two family patterns relevant in South Africa: nuclear and extended. This comes up because as Ziehl (2002) argues a nuclear household *structure* can exist within both a nuclear and an extended family *pattern*, meaning researchers should be cautious about counting up the share of nuclear households to draw conclusions about family patterns. She suggests instead that researchers look to the household structures unique to each pattern. These are couple and single-person households for the nuclear family pattern and extended households for the extended family pattern. The point is there is a subtle distinction between the question 'are family patterns changing?' and 'are household structures changing?' even if household structure can be used in part as evidence for both.

In the debate about households in South Africa, authors using either approach usually treated single-person households as 'nuclear' (Steyn, 1995; Amoateng et al., 2007; Ziehl, 2002). Amoateng et al. (2007) record a *drop* in their definition of nuclear households between the 1996 and 2001 censuses, but because single-person households (which the authors had separated out) increased, they wondered whether South African households were undergoing a process of 'nuclearisation'. In her assessment of family patterns, Ziehl (2002, p 45) argues that Black Africans are mainly following an extended family pattern because "...the single-person household is extremely rare among Black South Africans". Single-person households are very rare in the data Ziehl (2002) draws on which is mainly small localised studies from the early 90s. However, she also reports results from Simkins (1986) who uses the 1980 census data to find that 7% of rural and 6% of urban households were single-person.[13] Compare this to 2011, where we found that 27% of households were single-person.

How then to treat the single-person household given how its growth has revolutionised the household landscape in a way that was only emerging when this debate was first happening at the turn of the 21st century? The term 'nuclear family' was first coined by the family anthropologist Bronislaw Malinowski in the 1930s (Malinowski, 1930), when single-person households were practically unheard of (Snell, 2017). The original usage of the term and ensuing debates in the field of family anthropology were mainly concerned with the whether the nuclear family form was universal or essential for the nurturing of children and performing other essential economic, sexual and educational functions (Malinowski, 1930; Murdock, 1949; Hendrix, 1975). But, living alone is not only different to living with immediate family members as we discussed in the previous section, its new in a historical sense (Snell, 2017; Klinenberg, 2016). Single-person households only started occurring in meaningful numbers in the developed world after the 1960s and quickly started eclipsing households comprised of couples and two parents with children as marriage rates and fertility fell. This structural trend is today a well-recognised demographic phenomenon and lumping couple and two-parent households together with single-person households obscures this change. For example, if we were to apply the structure logic to Figure 8.1 we would easily conclude that households are becoming more nuclear. This happens *only* because single-person households increase so aggressively while other household structures that would be considered nuclear (like couples and couples with children) have either declined or remained stable at low levels. If we applied the family pattern logic, we would interpret the meteoric rise in single-person households amongst Black African men as meaning they were now subscribing to a nuclear family pattern. Both of these conclusions sit uneasily with us for a few reasons.

[13]The 1980 census data excludes the four independent homelands, but as Ziehl (2002) points out, these areas were predominantly rural and living alone can be expected to be rare.

Firstly, the group who are mainly responsible for the surge in living alone is Black African men who are historically associated with the system of oscillating labour migration (Wilson, 2001). Even though the living arrangements of migrants have become more permanent in the post-apartheid period, many still have strong ties to rural areas which embody cultural and religious ideas (Bank et al., 2020). Secondly, Section 8.1 described gendered patterns whereby men are most likely to form single-person households, and women are most likely to head complex households with children; but when not living alone, men are more likely to live with extended kin. Viewing these changes as a whole makes it hard to ignore that declining rates of marriage have interrupted family formation, so that the resultant rise in single-person households may not necessarily be due to more adherence to nuclear family ideals.

Thirdly, and as a result of both the previous two reasons, it seems likely that the single-person household might be increasingly mixing in with the extended family pattern, or representing an unconventional variant of the nuclear family pattern. According to Ziehl (2002), an extended family formation pattern might go as follows: a young woman lives in an extended family household; she gets married and moves into her husband's parent's household; husband and wife live there for a few years after having children; eventually she moves out with her husband and children and potentially some other kin to form a new household (if they move without kin or elderly kin pass away, the household looks nuclear, though). The cycle continues with her children. A nuclear family pattern goes as follows: a young man may leave the parental household to briefly live alone before marrying and living in a couple household; the household grows when they have children; but shrinks after the children leave; the household could be single-person again when a spouse dies.

In other words, the single-person household mainly finds expression in these family patterns as either a prelude to the couple household, or as a living arrangement of a widow or widower. This image is blurred firstly by labour migrants, but also by the rising share of never-married people. Whilst it is true that South Africans aged about 35 are most likely to be living alone, our work in the preceding chapters has shown that men, in particular, are increasingly living alone in later life e.g. ages 40-50. One reason for this appears to be the reduced chance that men will match up with a spouse. With cross-sectional data we cannot determine to what extent these are the same men, but one reading of the data is that for a growing share of men, the single-person household might be turning into effectively a life-long living arrangement. Based on the findings of Part II and III, a conceivable trajectory for Black African men in particular could be one that moves from the parental household, to living with other relatives as a young adult, to living in a single-person household (and possibly back with other relatives at advanced age, or not). In a time when martial rates are low, this could be an emerging trajectory for some portion of South Africans. This is neither a nuclear nor an extended family pattern, strictly-speaking.

The single-person household is nuclear to the extent that it represents a prelude to the formation of a nuclear family household on the road out of the parental household. But, to the extent that this is increasingly becoming a more permanent arrangement, increasingly linking up with extended households, or capturing labour migrants, we suggest that single-person households should be treated with caution and separated out. The trouble is there could very well be large shares of all of these types of single-person household (prelude types, migrant types, and permanent types) and it is impossible to tell them apart in a cross-section of data. This strongly motivates the need for panel data investigations of living arrangements over the life-course. In other words, just as Ziehl (2002) argued was the case for the classic nuclear household structure, so too it is becoming the case that the single-person households cannot be

used to reliably indicate change in people's underlying rules of how they form families (at least, in the cross-section).

We argue further that in an immediate structure-sense not living with anyone else separates people from those living in nuclear and extended household structures, which both ultimately encompass family ties of some sort and the ensuing differentiation between the latter two is about *which* ties (not that there are no ties at all).[14] The main point we are making is that in our present historical moment, single-person households are at the intersection of vital demographic, economic, and social change and tracking them in particular will help us identify and characterise that change. Separating out single-person households might go against the usual family patterns or even household structure categories, but without making this exception we run the risk of missing important changes that are extremely relevant in post-apartheid South Africa. Ultimately, more questions are raised by this discussion, some of which will only be answered as time unfolds and others by longitudinal or qualitative data.

8.3 Discussion: economic thinking as explaining change and raising more questions

So far in this chapter, we have addressed some questions of immediate interest regarding our findings about household formation. We naturally wanted to consider what our findings might mean for composition; address the increasingly important question of the meaning of so many more people living alone; and then reflect on what these ideas might mean for the main debate in the household literature. Having addressed these related, but somewhat distinct points, we now turn to pulling our ideas together and expressing them within the framework of economic thinking. A lot of material has been covered in this thesis, and ideas have been brought up and tied in from various disciplines. We now aim to use the discipline of economic thinking to help us cohesively describe our findings. This is not a definitive exercise, as ultimately, more questions are raised in the process.

8.3.1 Jobs, gender, and the reproduction of patterns of disadvantage

In Chapter 2, economic thinking provided us with three margins that might reduce household size, or speed up household formation: more income (lower costs), less need for household specialisation, and strengthening preferences for privacy and autonomy. Regarding the first margin, the legacy of apartheid spatial engineering might mean that moving (and potentially forming a new household in the process) is a necessary part of accessing income for many South Africans. Relatedly, marriage is tightly woven into the process of accessing income for women in a way that is not the same for men, as we have emphasised throughout this thesis. Women's historic disadvantage in the labour market and specialisation in household production made them reliant on male earnings for economic survival, often accessed via marriage. Excessive levels of male unemployment in the post-apartheid period have meant that women have had to adapt how they access income, a factor likely leading to broad-based change in marital rates and household formation.

[14] Another household type we haven't discussed is multi-person households comprised of only non-related members. Amoateng et al. (2007) separate out this category and it is small at less than 5% by 2001.

The second margin is specialisation, which relates to the need for specialisation within the household to produce household public goods and services. Specifically, this thesis has dealt with household gender specialisation, described in detail in Chapter 2, although there are other factors that might reduce the need for household specialisation we have not covered (e.g. technology, State-provided services). Household gender specialisation in South Africa has partly been reduced in the sense that women expanded the rate at which they participated in the labour market in the 1990s (Casale and Posel, 2002). This better enabled them to adjust when high male unemployment meant marriage could less and less provide them with their traditional access to income. A question we raised right at the beginning of this thesis was, why would South Africans be forming households so quickly in the face of high unemployment? Counter-intuitively, the answer may partly be precisely because unemployment is so high, eroding a key structure maintaining household gender specialisation.

However, household gender specialisation remains in other forms in South Africa to the extent that men continue to be unlikely to carry out household production (Hatch and Posel, 2018). The result is that the reduction in household gender specialisation that has occurred in the country has mainly worked to set up a double burden of labour and household production for women. This is reflected in the household compositional patterns documented in Section 8.1 of this chapter. Four planks reinforce this process in South Africa's context: firstly, prevailing gender norms that ascribe childcare to women and provision to men (Makusha et al., 2019; Timol et al., 2019), interact disastrously with the second point of open unemployment, undermining both men's marriageability (Posel and Casale, 2013) and the chance he will remain in the same household as his child. Thirdly, continued female disadvantage in the labour market undermines her ability to support herself (Casale et al., forthcoming). Fourthly, this situation is not alleviated by any State-supplied day care or other family policies, meaning women have to draw on private security nets of usually female relatives to cope with the double demand of work and child-care (Posel and Hall, forthcoming).

8.3.2 Competing explanations: raising more questions

The third margin coming out of economic thinking, is preferences. Embedded in the earlier question about whether households are becoming more nuclear, is a question about changing preferences. Questions about value-change are especially relevant based on our finding in Chapter 6 that *behavioural* change was behind more household formation. Household change reflects broader social transitions, connecting the work in this thesis to deeper questions. Answering these questions comprehensively is beyond the scope of this thesis. Our aim here is mainly to raise and clarify questions, as a way to emphasise that household change is tied up in these overarching processes. Preference change relates to our description in Chapter 2 of how the delay and decline in marriage and rise in cohabitation in the West has partly been attributed to ideational change related to the SDT, where people increasingly started to favour the Individual over the Family (Lesthaeghe, 2010). In South Africa, Russell (2004, p 65), for example, advances that a social transition is clearly underway with implications for both extended and nuclear family traditions, and that urban Black Africans are "evenly divided" in their adherence to either the norms of the past or their enthusiasm for Western individualism.

However, taking household compositional change as a whole, so clearly shaped by the decline in marriage, makes us cautious about interpreting the growth of single-person households amongst Black African

men as subscription to nuclear family ideals. The SDT value-change explanation is often contrasted with an alternative explanation called the 'pattern of disadvantage' (POD), where people are adopting more ad-hoc and less permanent living arrangements as a coping mechanism in today's precarious economic environment (Lesthaeghe, 2020). A third explanation comes from convergence theory, also described in Chapter 2, which is the theory that as economies modernise, smaller household sizes make more sense and are more feasible. A fourth factor is, of course, the role played by South Africa's unique history. These competing explanations raise interesting questions such as, to what extent are we observing more household formation because people are acting on new social aspirations and priorities, and to what extent are we are seeing vulnerable people adapting to precarious economic realities? To what extent are households getting smaller because of the gravitational pull of modernisation, compared to the unfolding legacy of apartheid spatial and social engineering? To what extent are we observing the most destructive effects of labour migration on family formation taken to its extreme conclusion, whilst the extended kin system adapts, yet again, to cope with the economic circumstances of the post-apartheid period?

Most probably, all of these forces - value change, patterns of disadvantage, economic modernisation, and, South Africa's history - are playing some part, although we speculate that the POD explanation in combination with unfolding historical legacies are likely playing the dominant role. We know that the labour market remains both the most important source of household income in South Africa and highly exclusive at the same time (Woolard and Klasen, 2005; Wittenberg, 2017). Chapter 2 described how South Africans entered the post-apartheid period with already-high levels of poverty, unemployment, and inequality, and how these levels have not budged in meaningful ways for the majority of people. Attitudinal data also shows that South Africans view marriage positively (Mohlabane et al., 2019), meaning high male unemployment is likely the prime candidate behind the decline in rates of marriage (Posel and Casale, 2013), as opposed to broad-based change in people's ideals about marriage. Craigie et al. (2018) provide some empirical support for the idea that in a high-unemployment setting, women's marital prospects adjust downwards and their household formation behaviour accelerates accordingly, albeit from a setting in the US. In qualitative work from South Africa, migrants have described the long-distance travel they embark on for work as a stark trade-off and voiced the feeling of having no choice in the matter (Hall and Posel, 2019). Although many people still migrate to urban areas to work, they remain 'double-rooted', maintaining strong ties with rural areas as sites of culture and tradition (Bank et al., 2020).

To some extent, all four of these explanations are likely converging and value-change and economic realities are interacting endogenously. Relevant in this regard is Honwana (2012)'s concept of 'waithood' - that South African youth today are suspended in an interstitial period between childhood and the onset of adulthood. Young adults are too old to be considered children, but constrained by their socio-economic circumstances to attain the usual markers of adulthood, like getting a job and getting married. Waithood is frustrated by the influence of Western consumer culture and ideals that are increasing accessible to youth with internet and cellphone technology. Honwana (2012) argues that in this milieu, youth are finding new ways to gain social status and meet their social, economic and sexual needs that might not conform to the ideals and norms of their parent's generation, e.g. by leveraging transactional sex (Hunter, 2002). Regardless of whether social change in South Africa is mainly based on value-change or patterns of disadvantage, the key insight from Esping-Andersen and Billari (2015) is that the result for women will still be the double burden of work and child-care unless gender relations, themselves, change in structural

ways.

8.4 Conclusions so far

The division of labour in the household structures the relations between men and women, and their livelihoods. The change in household formation observed over the post-apartheid period directly reflects how these aspects have changed in the presence of high levels of unemployment and falling rates of marriage. The consequence has been gendered household formation patterns where men mainly form single-person households, whilst women mainly form complex households with children; and, more South Africans are living in households with other relatives. This growth at the poles of household complexity, so clearly influenced by declining marital rates, makes it hard to simply conclude that households in South Africa are becoming more nuclear. As men are less and less likely to be found in households including women, the implication for the livelihood strategies of these households is that they become increasingly reliant on grant income and female labour market earnings. The result is that the household formation and structure patterns we are seeing continue to reproduce the double burden for women of balancing both earning an income whilst caring for children. For men, it is not categorically true that because they are more likely to be living alone, they are better-off economically. Although they might not have to share their income with anyone else, they also have no one else with whom to share costs (as far as the survey can see). More people living alone may also raise concerns for public health given the connection made in the literature between living alone and scoring highly on dimensions of loneliness or social isolation, which in turn have been shown to pose severe risks to mental and physical health.

Chapter 9

Conclusion

9.1 Households in social transition

Who you know and how much they have matters enormously for almost every social outcome of interest for the post-apartheid research agenda. In the nationally-representative data available to researchers in South Africa, the household is often the only way social connections are captured. Who makes up the household and who does what in terms of labour market or household production; as well as, who can draw on ties outside of the household is critical to the day-to-day welfare and well-being of those who live in that household. Child health and progress through school; gender-based violence and continued female disadvantage in the labour market; family formation and livelihood strategies - these are just a few social outcomes that are inextricably bound up with the household. This means that the progress of South Africa's post-apartheid project cannot be properly understood or contextualised without a solid understanding of the accompanying household change. And, households have indeed changed in a number of critical dimensions since the first democratic election in 1994. One of the main ways in which this has happened is that household growth has outstripped population growth almost two to one. This thesis provides the most thorough investigation of country-level trends in post-apartheid household *formation*, so far - an aspect of household change relatively less-studied.

9.2 What have we learned about household formation?

Between 1995 and 2011, South Africans formed more and smaller households and who was living in those households changed in systematic ways. This change was not only the outcome of population growth and falling fertility; it was also a function of people splitting off more quickly from existing households and forming new ones. Our research has underscored that the continuing decline in the marriage rate is a core driver of this process. Whilst in the developed world, an important role has also been played by

a growing population share of the elderly and higher divorce rates (Bradbury et al., 2014; Peichl et al., 2012), this is not the case in South Africa. Instead, the story in South Africa is about people not getting married in the first place, and how South Africans have adapted existing livelihood structures (e.g. labour migration) and support systems (e.g. extended kin networks) to cope with the fallout. Change in marital rates have had profound effects on household formation because marriage is so closely bound up with women's economic welfare, in a way that is not quite the same for men. Historic female disadvantage in the labour market and specialisation in household production meant women were dependent on men's earnings, often accessed through marriage. However, as male unemployment shot up towards the end of apartheid, marital rates consistently declined as his marriageability suffered. The result has been change in household formation that reflects gendered access to labour market and social grant income across the life-cycle.

Female household formation surged immediately after apartheid, potentially in response to an expansion of women's rights by a series of legislative changes that happened in the 1990s. These included more protection against domestic abuse and more reproductive rights; for Black African women, the new constitution also provided women with the right to own their own property which they had been denied under apartheid. These new freedoms, in combination with more women entering the labour market, encouraged and enabled more women to form their own households. However, as the decade of the 2000s wore on, it became clear that the labour market would remain highly exclusive and without meaningful family policies from the State, the ability of many women to participate in the labour market was limited by childcare responsibilities. Female household formation plateaued in the 2000s, falling in step with the growth of the female adult population.

The household formation of Black African men remained relatively stable immediately after apartheid, while all other groups of men reduced their household formation, most notably white men. Then in the 2000s, the rate at which Black African men formed households started to gain momentum. Behind this aggregate picture, Black African men astronomically increased the rate at which they lived alone. The number of people living alone increased by 150% from 1.58 million in 1995 to 3.98 million 2011; of whom, 2.4 million (or 60%) were Black African men in the later year. Developed countries are the vanguard of living alone, but South Africa has rates of single-person households that are comparable to the United States and New Zealand owing in part to its long history of migrant labour (Ortiz-Ospina, 2019). An influential driver behind the rise of the single-person household in South Africa appears to be declining rates of marriage; which also started transforming the profile of female heads, even if by the 2000s their overall rate of household formation had steadied.

We find, for example, that young women are highly reliant on marriage as a route out of the parental home because they are a demographic facing one of the highest unemployment rates. As marital rates have declined over the post-apartheid period, this meant young women in 2011 left home at about the same rate as those in 1995, and those in rural areas were more likely to remain at home. For the growing share of women who entered middle-age without having formed a union, this was the time when they become increasingly likely to form their own household. This happens now because, starting at about age 30, women enter the years of their life when they are most likely to have a job, making the move financially feasible without a spouse. Based on a mechanism for which evidence exists for the US (Craigie et al., 2018), we suspect women were also more likely to make the move in 2011 compared to 1995 because they knew their chances of marriage were lower, making them less willing to wait around

for a spouse with whom to form a household. Many women would already have had a child by this point and might have moved out with other probably female relatives to form a complex household. Women in older generations who faced better marital prospects in their youth, often found themselves heading households when they were widowed in later life.

Men, on the other hand, have more options for forming households than women: getting a job, getting married, or both. Young white men delayed leaving home between 1995 and 2011; but, young Black African men started leaving home sooner, around age 25 by 2011 and probably for work-related reasons. It was less likely a young man would be married by this age in 2011 compared to 1995, making him more likely to live with other relatives or live alone when he leaves home. For men who later did form a union, this almost guarantees headship across the life-cycle: in 2009-11, over 80% of married-employed men over the age of 35 were household heads; and, 60% of men over the age of 60 were married household heads. However, a growing portion of men are becoming lifetime bachelors and, although we cannot say with certainty using cross-sectional data, this might mean that more and more men simply continue to live alone as a result as they age up.

This process has immediate consequences for livelihood strategies and a broad sweep of other social outcomes. Two main outcomes are, firstly, that many more people are living alone, mostly men, with currently ambiguous implications for their economic welfare and physical and mental well-being. Secondly, gendered household formation patterns are very likely reinforcing the double burden of labour market and household production for women. Without a reduction in the unemployment rate, more progressive gender attitudes towards child care and fatherhood, and, State family policy that might alleviate this burden (e.g. day care), these patterns will probably only serve to entrench female disadvantage in South African society. We return to these points shortly. First, though, we reflect on how we got there by reviewing the three main contributions made in this thesis: measurement (especially of small households); the role of marriage; and, home-leaving. We review each of these three contributions in turn by first describing our approach, how it represented a contribution to the existing literature, what any limitations might be, and whether this aspect uncovered any new questions for future research.

9.3 Measuring household change better in the OHS and GHS

We were interested in what we could learn from survey data about how household formation has changed over the post-apartheid period. As such, we constructed the OHS-GHS series, a harmonised series of twenty nationally-representative cross-sectional household surveys from StatsSA, collected on an annual basis. The series originally comprised the October Household Surveys (1994-9) and most up-to-date versions of the General Household Surveys (2002-2015) at the time, covering the period 1994-2015. However, we faced a serious data quality impediment in the form of the survey weights released with the data. StatsSA calibrate the weights for people and households independently (StatsSA, 2010*c*), meaning researchers can use the same sample of people and households, but get different answers to the question 'how many households are there in South Africa in a given year?' just by applying a different weight. This undermines the extent to which the data is fit for the purpose of studying households. Indeed, this practise undermines the extent to which the data is fit for the purpose of studying any question requiring people to relate sensibly to the households in which they live. For our purposes in particular, it

compromised our ability to plot trends in the number of households, and added unnecessary complexity to calculating household headship rates.

Part I then comprised our efforts to improve the quality of the data before we could embark on more substantive analysis. In Chapter 3, we used cross-entropy estimation to reweight the data to solve the statistical and conceptual problems created by what we called the StatsSA dual-weight system. In the process, we had to restrict our time period to the fifteen surveys between 1995 and 2011. In Chapter 4, we went a step further and paid attention to sampling and benchmarking issues to improve the overall quality of the calibration. The outcome is a series of integrated cross-entropy weights that uses both person and household auxiliary information; coherently links people with the households in which they live; produces benchmarked trends in the total number of people and households in each year; and which specifically provides accurate counts of small households.

Measurement is an important theme in the literature on households in South Africa. Partly, this is because defining, and therefore measuring, the household is difficult. Researchers from various disciplines approach households from alternative points of view, and might emphasise a slightly different aspect or have slightly different priorities about how to define a household depending on their research question. However, most researchers still turn to household survey data to gather evidence for questions falling broadly under the umbrella of 'what is happening to households?'. More specifically, they often turn to the household survey data released by StatsSA when wanting to answer this question in a way that is representative of the country as a whole. Therefore, the quality of these data matters enormously.

Measurement is also a theme in this literature because data quality or lack of comparability between sources has hampered the ability to study households. Early debates about households in South Africa were hindered by the inability to make comparisons across data sources (Ziehl, 2001). Many researchers drew on small highly-localised studies, some of which differed in their definition of a household (Seekings, 2008). Even today, when we have more than 20-years worth of StatsSA household surveys with an invariant definition of a household, problems remain, including those created by the dual-weight system. Accurate and reliable data sources are a prerequisite for researchers and policymakers concerned with households. Contributing to the work of improving the data quality of our national household surveys represents a main contribution of this thesis. For the first time, we have extracted a series of total household counts benchmarked using both person and household information for almost every year in the period 1995-2011.

We focused, in particular, on providing accurate counts of small households. Most country-level research on single-person households in South Africa relies on the three censuses only because of the known problems with the undersampling of small households in these surveys (Kerr and Wittenberg, 2015). By specially benchmarking on the number of one- and two-person households, we were able to reliably use the surveys (and not just the censuses) to study the important sub-group of single-person households at the country-level. This allowed us to fill in the gaps between the three census points and use information from multiple years to inform us about change over time in more detail, e.g. age-profiles, marital status. Trends are key sense-checking tools. We can be more certain a change is real if we observe smooth trends that move realistically over time, rather than big jumps, which we would not be able to detect if we only see a few isolated data points. Essentially, more information is better, and fifteen data points is better than three data points.

Measurement of households in the OHS-GHS series is not without limitations. Some of the biggest

limitations for this thesis arise from the restrictive way in which households are measured in StatsSA surveys. The most important aspect of household formation that we could not study in these data was the role of labour migration. The StatsSA data do not record absent household members, or collect detailed-enough information about labour migrants and remittance flows. As such, we had to tie in what we know about labour migration extracted from other sources by other authors. It is very difficult, if not impossible, to answer the very important question of to what extent household proliferation has been driven by changing patterns of labour migration with these data. Labour migration also overlaps with the critical demographic process of urbanisation, also connected to shifts towards smaller living arrangements and less marriage. Urban location variables exist in the OHS-GHS data, but the distinction between what space counts as urban or rural has changed over time, raising questions about the comparability of these variables over time.

The second major omission in the survey relates to parent-child relationships. Parenthood is only observable in these surveys to the extent that parents and children live in the same household. This is often not the case in South Africa because of high levels of internal labour migration and strategies to maximise schooling that cross large distances (Hatch and Posel, 2018; Posel and Hall, forthcoming). In some early versions of the OHS and GHS, women were asked about their fertility history, but there is no way to tell if men are parents if they are not co-residing with their children. In the current set-up of the StatsSA surveys, we have very little information about how family ties might transcend the household.

Despite these drawbacks, the work in this thesis has shown, firstly, that it is possible to solve the problems created by the StatsSA dual-weight system and therefore restore some important quality attributes to the data. Secondly, we have undertaken important work in uncovering and mitigating some of the issues arising from the undersampling of small households. The work on reweighting the OHS-GHS series is not complete. There is more to be done in terms of fine-tuning the weights and dealing with some of the unavoidable data quality issues with the benchmark sources themselves. However, the work so far has made some significant progress towards improving the data quality of our national household surveys. These surveys are used not only by researchers from a broad range of fields, but also by government as the basis of evidence-based policy-making, planning, and monitoring.

9.4 Understanding national-level trends in household formation: incorporating marriage

The data set up in Part I formed the basis for analysis of household formation in Parts II and III. There is no single definition of household formation in the same way as there is no single definition of a household. This brings forth the question of measurement and makes it a good strategy to approach the subject from different angles and consolidate conclusions across approaches. This is what we do over the course of Parts II and III. In Part II, we begin by investigating the aggregate change: tracking the process by which the number of households in survey data expanded from 9.1 million in 1995 to 14.5 million in 2011. The proliferation of the number of households was one of the most important shifts in the household landscape over the post-apartheid period. Many researchers have noted the ensuing drop in household size, but only a handful of studies have tried to explain this phenomenon, and mainly using localised data (Wittenberg and Collinson, 2020; Wittenberg et al., 2017).

To study household proliferation, we focused on household heads because, by definition, they tally with the number of households in our data. This is slightly controversial because many scholars have rightly questioned the meaning of the household head (Budlender, 2003; Presser, 1998), which is ultimately an artefact of survey enumeration. The point, though, is that household headship is a good proxy to capture the person who in most cases made the decision to form the household. In fact, we would argue this is probably more true in South Africa, than a country like the US, for example, because South African society remains highly patriarchal, making the idea of household authority a familiar one (Timol et al., 2019; Rogan, 2016).

The benefit of this approach is that it allowed us to track the patterns behind South Africa's proliferation of households directly. Chapter 5 comprised a thorough description of trends in household headship and living alone, along with how these cohered with age, labour market and grant income access, and marital status. As far as we are aware, these trends have not been systematically studied in the way we have done so here, or with the benefit of the careful reweighting work of Part I. New insights include the description of trends in the opening section of this conclusion: the boom-plateau pattern of female household formation and the slow-boil pattern for men. Other important findings were that women who were prime-aged and older specifically started ramping up their rate of household formation; whereas, first young men, but later also older men started to live alone more often.

An important story that started taking shape in Chapter 5 was the role of the marriage market and how this interacted in gendered ways with the labour market. One of the contributions of this thesis has been including the marriage market more cohesively within a framework of household formation in South Africa. Marriage is a well-established driver of household change in the West and East Asia, but the econometric literature on the topic in South Africa has focused on the labour market, omitting marital status even as a control variable, because of its endogeneity with employment. Chapter 6 sought to more rigorously develop the ideas of Chapter 5 by testing the relative importance of different drivers of household formation in a *ceteris paribus* setting. We set up a model and paid attention to modelling this endogeneity by including an interaction between marital and labour market status. As we mentioned above, part of the reason why marital status has proved so influential is precisely because it is so closely bound up with women's economic welfare in a way that is different to men. The interaction between marital status and labour market status proved much more important for women; whereas marital and employment status each yielded larger separate effects of their own for male headship. Marriage worked in opposite directions for male and female headship: it made men more likely to be heads, but women, less likely. However, both men and women were significantly more likely to live alone if they were not married.

As it became clearer that poorer marital prospects were playing an influential role in household formation, it raised the question of how this was happening. Specifically, we were interested in whether we were seeing more households simply because there were more never-married people, or because the *rate* at which never-married people were forming households increased. To answer this question, we decomposed our models of household headship and living alone across time using an Oaxaca-Blinder decomposition. This application was complicated by our outcome variables being binary. Essentially, what might have been a logical model choice for the regression model, implied a less-optimal decomposition choice. But, the optimal modelling choice for the decomposition, required a potentially less-optimal model for the regression. We therefore ran different versions of both the model and the decomposition and our results

were highly robust. Acceleration in the rate at which never-married people, in particular, were forming households was a key contributor to the increase in the headship rate and the rate at which more people were living alone. That this result emphasises *behavioural* change, raises more questions: why would never-married people form households so much more quickly? We do not go so far as to answer this question in this thesis, and leave it for future research. We do, however, expect never-married people are probably forming households more quickly because they know they are less likely to marry, providing less reason to wait around to form a union before moving out. Some empirical evidence that such a mechanism could exist when unemployment is high comes from the US (Craigie et al., 2018).

A limitation of the work in Part II is that it remains descriptive. Partly, this is because our data is cross-sectional. For example, in Chapter 5, structural breaks in the headship trend coincided with changes in the underlying instrument from the OHS years to the GHS years. In our analysis of age, we also found first more middle-aged and later more senior men started living alone as time progressed. However, we could not verify if these were the same individuals or test the veracity of the trends using the same sample of people. Instead, we relied on analysis of birth cohorts to deal with this. The model and decomposition are also only descriptive. This is also partly a function of the nature of the subject matter. Household formation is an intrinsically complex social process, complicated at every step from the definition of a household to the endogeneity of tightly interrelated long-running social processes like population aging, marriage, employment, value change, and family formation (Bongaarts, 2001). Many of these processes are so closely interconnected that studying households using time-series cross-sectional data, like we do here, will rarely meet the standards of causal analysis in applied economics today. Instead what is best, is good-quality descriptive work, again bringing us back to the importance of reliable measurement.

Furthermore, the cross-sectional surveys making up the OHS-GHS series are the only nationally-representative data sources available for the period immediately after apartheid. Therefore, providing a description of country-level trends in household formation since the beginning of the post-apartheid period is one of the contributions of Part II. Related to this, Part II provided some of the most detailed descriptions of living alone in South Africa, so far, made possible by the data contribution of Part I. A major contribution from Part II was developing in more detail the role worsening marital prospects have played in household formation in the post-apartheid period, with a keen eye on how this interacted with gendered access to the labour market. This has opened up more questions about the behaviour of never-married people; as well as, highlighted the need for a better understanding of how the labour market, marriage market, and household formation all interconnect.

9.5 A first description of trends in home-leaving in South Africa

The second way in which we study household formation is the process of leaving the parental home in Part III. The process by which young adults leave the parental home is a natural focal point in the literature about household formation. Leaving home coincides with other important first events in young adulthood, like first union; and is associated with the establishment of a new household in the nuclear family pattern. Very little macro-level empirical work has been conducted on home-leaving in South Africa, and certainly no studies on trends. This meant that in Chapter 7 in Part III, we answered some primary questions about home-leaving for the first time; such as, how age of home-leaving has

changed over time; what types of households home-leavers live in; and, how home-leavers look different to home-stayers and adults at other parts of the life-course.

We found that by 2011, 25 was the age by which most young adults in South Africa had left home. This was, however, the outcome of divergent patterns by race due to different groups in South Africa adhering to diverse family patterns and also facing various socio-economic circumstances. We showed that whites have delayed their rate of leaving home, whilst Black African men, in particular, have started leaving home sooner. This connected well with our earlier finding that more young men are living alone. Black African women followed different patterns by urban-rural location, with urban women leaving sooner, but rural women becoming more likely to still be at home by 2011. Home-leavers emerged as particularly dependent on labour market income in the absence of social grants targeted, in particular, at the large unemployed young adult population. Since young women face higher unemployment rates than young men, women were notably more reliant on marriage as a channel out of the parental home. These patterns were remarkably similar to those of Southern European countries, where home-leaving happens in the mid-20s, marriage is also definitive for home-leaving, and State welfare spending favours the aged over the unemployed (Billari and Wilson, 2001).

This chapter was limited by the data being cross-sectional. Panel data is particularly well-suited to the study of home-leaving which is intrinsically a question about trajectories over the life-course. Cross-sectional data also precluded our ability to identify returns to the parental household, for example. Panel data represents a fertile avenue for future research on this topic, for which many new questions have been raised by this initial work. Nevertheless, the profile of home-leaving in Chapter 7 provides answers to some first-order questions about this process in South Africa for the first time, and as we mentioned above, these data are the only source that could answer these questions for the early post-apartheid period. The findings in Part III strengthened the conclusions from Part II and placed these developments more robustly within the life-cycle perspective.

9.6 Implications of our findings: questions and connections to broader social transitions

In Chapter 8, we investigated the implications of our findings about changing household formation. We found that household composition has grown at the poles of complexity: more people are living alone and more people are living with extended kin. Furthermore, these patterns diverge sharply by gender, with men mainly forming single-person households and women mainly forming complex households, usually including children and other female relatives. Households headed by never-married women have even become *more* likely to include children under the age of 15 over time despite the downward pull of falling fertility rates. This process overall is very likely reinforcing women's double burden of work and childcare. A shortcoming of these data is we cannot investigate in detail how remittances are flowing between households, meaning we cannot tell to what extent fathers or other male relatives living alone are remitting to these female-headed households. However, even if he is sending money for the upkeep of some individual household members (Posel, 2010; Madhavan et al., 2008), it is unlikely that this would be completely alleviating her need to work in the majority of cases (Hatch and Posel, 2018). Several planks reinforce this situation, including female disadvantage in the labour market undermining her ability to

attain financial independence; high male unemployment destabilising the marriage market; gender norms that ascribe childcare to women and economic provision to fathers; and, a lack of family policies from the State (e.g. day care, paternity leave) that could alleviate her double burden.

We also discussed how, at present, we are unsure of the exact implications of having so many more South Africans living alone when the full spectrum of social and economic factors are taken into account. For example, although single-person households might appear better-off than other household types in per capita income terms, living alone is relatively more expensive from a cost point of view (Lanjouw and Ravallion, 1995). Studies from the developed world have associated living alone with more self-determination and freedom (Jamieson et al., 2009); however, living alone has also been linked to a higher chance of loneliness and social isolation, in turn related to poorer physical and mental health outcomes (Holt-Lunstad et al., 2010; Pantell et al., 2013). The dynamics surrounding living alone are further complicated by South Africa's history of labour migration and stretched households. At least some questions about the well-being of people living alone are empirical ones that could be answered with data that exists for South Africa.

The recognition that living alone is critically and ambiguously caught up in contemporary social change informed how we interpreted the meaning of our findings for the earlier debate in the literature about whether households were becoming more nuclear in South Africa. The mushrooming of single-person households have transformed the living arrangements of South Africans in a way that was not detectable when researchers were most actively debating whether households were becoming more nuclear. This was not detectable either because researchers were working in early years when levels of living alone were still low; or, for researchers working later, because quality issues in the household surveys meant they were not accurately reflecting the level of living alone. The answer to whether South African households are becoming more nuclear depends very much on whether one considers the many more single-person households, mainly occupied by Black African men, as nuclear. Certainly, there are many more nuclear household *structures* in the sense that there are many more single-person households, but this contrasts with the simultaneous increase in the share of extended households. Whether more living alone has coincided with a concomitant change in people's underlying ideals about family formation is much less clear. As such, we cautioned against uncritically treating the single-person household as nuclear and our contribution to this original debate is thus mainly a cautionary note about how to interpret the rise of solo-living, which is at the nexus of local historical legacies and critical social processes of our time.

In Chapter 8 we also brought more structure to bear on our findings about household formation using the economic thinking set up in Chapter 2. The main insight, already brought up earlier in this conclusion, was that both income access and household specialisation are profoundly wrapped up with gender, explaining why marriage - as an institution governing both gender norms and household formation - has emerged as such a vital driver in this thesis. Finally, in this section we also connected household change to broader social processes and raised questions about what the underlying social drivers might be, especially relevant since we found that household formation *behaviour* was so pivotal. For example, are people changing their household formation behaviour because their values are orienting towards more privacy, more individualism, and more nuclear family patterns? Or, are people adjusting their household formation to cope with precarious economic conditions of the post-apartheid period? Relatedly, to what extent are we observing new social aspirations and the pull of the future, and to what extent are we viewing the unfolding of the past? Are we observing agency or adaptation? Declining rates of marriage

in the developed world have been linked to the ideational change of the Second Demographic Transition and women entering the labour market (Lesthaeghe, 2010). No doubt value and ideational change is playing some part in social change in South Africa today (Russell, 2003; Honwana, 2012). However, this change also cannot be de-linked from high open male unemployment and the unfolding legacy of labour migration and apartheid social and spatial engineering on family formation in this country.

9.7 In conclusion: a richer understanding of household formation in South Africa

The patterns and drivers of household formation revealed in this thesis broadly echo those found in the international literature on household formation in terms of age, marriage and employment. However, factors specific to South Africa have distinctly shaped the process in this country. These include levels of unemployment unseen in the developed world; a long history of labour migration; the spatial engineering of apartheid; and welfare spending that prioritises children and the aged over young unemployed people and mothers who need to work. Thus, the pattern in South Africa is less driven by higher divorce rates or aging populations as in the West and East Asia, and much more by the fact that South Africans are less likely to marry in the first place, related in turn to high levels of male unemployment.

The major insight from this thesis is that household formation in South Africa has responded to how high male unemployment has washed away a key plank supporting gendered household specialisation. As women could less and less rely on marriage as a means to access income - and likely also more proactively in combination with new freedoms afforded by the onset of the new democratic era - they adapted their household formation behaviour. The resultant decline in union formation has similarly prompted a robust rise in the rate at which men live alone. Circumstances in labour markets and marriage markets are most closely woven together and reflected through the household. Change in one of these aspects will have profound implications for the others, and often for a range of other social indicators, highlighting how households represent part of the most critical context for the post-apartheid period.

Most recently, the centrality of the household to people's livelihoods has been emphasised by the drastic changes in people's co-residency that happened in response to the Covid-19 lockdowns in 2020 (Posel and Casale, 2020). The contribution of this thesis towards our understanding of how households adapted over the course of the post-apartheid period provides important insight into major constraints and opportunities to advance the post-apartheid project; as well as, how we can expect South Africans to adjust to the economic fallout of the Covid-19 pandemic going forward. Our description of how household formation may be reinforcing women's double burden presents clear constraints on women's ability to improve their welfare with direct implications for the children being raised in their households. Single-person households represent a rich area for future research with many open questions about their welfare, their motivation for living alone, and their social connections outside their own household. Overall, household change described in this thesis cannot be de-linked from gender relations and the myriad ways in which the unfolding legacy of South Africa's past continues to create sources of vulnerability in people's everyday lives.

Bibliography

Aassve, A., Billari, F. C., Mazzuco, S. and Ongaro, F. (2002), Leaving home: A comparative analysis of ECHP data, *Journal of European Social Policy* **12**(4), 259–275.

Abrahams, N., Jewkes, R., Martin, L. J., Mathews, S., Vetten, L. and Lombard, C. (2009), Mortality of women from intimate partner violence in South Africa: a national epidemiological study, *Violence and Victims* **24**(4), 546–556.

Aggeborn, L. and Nyman, P. (2020), Intergenerational transmission of party affiliation within political families, *Political Behavior* pp. 1–23.

Albertyn, C. (2009), 'The stubborn persistence of patriarchy'? Gender equality and cultural diversity in South Africa, *Constitutional Court Review* **2**(1), 165–208.

Althaus, F. A. (2000), Work in progress: the expansion of access to abortion services in South Africa following legalization, *International Family Planning Perspectives* **26**(2), 84–86.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E. and Zucman, G. (2018), *World inequality report 2018*, World Inequality Lab, Paris School of Economics.

Ambler, K. (2016), Bargaining with grandma: The impact of the South African pension on household decision-making, *Journal of Human Resources* **51**(4), 900–932.

Amoateng, A. Y. (1997), The Structure of Urban Black Households: New Survey Evidence from a Coloured and an African Community on the Cape Flats in the Western Cape of South Africa, *African Sociological Review* **1**(2), 22–40.

Amoateng, A. Y., Heaton, T. B. and Kalule-Sabiti, I. (2007), Living arrangements in South Africa, *in* 'Families and households in post-apartheid South Africa: Socio-demographic perspectives', HSRC Press Cape Town, pp. 43–59.

Andres, L., Anisef, P., Krahn, H., Looker, D. and Thiessen, V. (1999), The persistence of social structure: Cohort, class and gender effects on the occupational aspirations and expectations of Canadian youth, *Journal of Youth Studies* **2**(3), 261–282.

Ardington, C., Case, A. and Hosegood, V. (2009), Labor supply responses to large social transfers: Longitudinal evidence from South Africa, *American Economic Journal: Applied Economics* **1**(1), 22–48.

Awumbila, M., Teye, J. K. and Yaro, J. A. (2017), Social networks, migration trajectories and livelihood strategies of migrant domestic and construction workers in Accra, Ghana, *Journal of Asian and African Studies* **52**(7), 982–996.

Baland, J.-M. and Ziparo, R. (2017), Intra-household bargaining in poor countries, WIDER Working Paper 2017/108, United Nations World Institute for Development Economics Research.

Bank, L. (2020), Displaced urbanism: City shack life and the citizenship of the suburban house in the rural Transkei, *in* 'Migrant labour after apartheid: the inside story.', HSRC Press, Cape Town, chapter 16.

Bank, L., Posel, D. and Wilson, F. (2020), Introduction: Migrant labour after apartheid, *in* 'Migrant labour after apartheid: the inside story.', HSRC Press, Cape Town, chapter 1.

Bartus, T. (2006), Marginal effects and extending the Blinder–Oaxaca decomposition for nonlinear models, Stata Users Group. United Kingdom Stata Users' Group Meeting 2006 05.

Bauer, T. K. and Sinning, M. (2008), An extension of the Blinder–Oaxaca decomposition to nonlinear models, *AStA Advances in Statistical Analysis* **92**(2), 197–206.

Becker, G. S. (1960), An economic analysis of fertility, demographic and economic change in developed countries, *National Bureau Committee for Economic Research* **209**.

Becker, G. S. (1965), A theory of the allocation of time, *The Economic journal* pp. 493–517.

Becker, G. S. (1981), *A Treatise on the Family*, Harvard university press.

Bentley, K. (2004), Women's human rights & the feminisation of poverty in South Africa, *Review of African Political Economy* **31**(100), 247–261.

Bhorat, H. and van der Westhuizen, C. (2012), Poverty, inequality, and the nature of economic growth in South Africa, Working Paper 12/151, Development Policy Research Unit, University of Cape Town.

Biewen, M. (2012), Additive decompositions with interaction effects, IZA Discussion Paper No. 6730.

Billari, F. (2004), Becoming an adult in Europe: A macro (/micro)-demographic perspective, *Demographic Research* **3**, 15–44.

Billari, F. C., Philipov, D. and Baizán, P. (2001), Leaving home in Europe: The experience of cohorts born around 1960, *International Journal of Population Geography* **7**(5), 339–356.

Billari, F. C. and Wilson, C. (2001), Convergence towards diversity? Cohort dynamics in the transition to adulthood in contemporary Western Europe, Max Planck Institute for Demographic Research, Working Paper No. 2001-039.

Blinder, A. S. (1973), Wage discrimination: reduced form and structural estimates, *Journal of Human Resources* pp. 436–455.

Blumenstock, J. E., Chi, G. and Tan, X. (2019), Migration and the value of social networks, CEPR Discussion Paper No. DP13611, Centre for Economic Policy Research, United Kingdom.

Bongaarts, J. (2001), Household size and composition in the developing world in the 1990s, *Population Studies* **55**(3), 263–279.

Boonzaier, F. (2005), Woman abuse in South Africa: A brief contextual analysis, *Feminism & Psychology* **15**(1), 99–103.

Bradbury, M., Peterson, M. N. and Liu, J. (2014), Long-term dynamics of household size and their environmental implications, *Population and Environment* **36**(1), 73–84.

Branson, N. and Wittenberg, M. (2014), Reweighting South African national household survey data to create a consistent series over time: a cross-entropy estimation approach, *South African Journal of Economics* **82**(1), 19–38.

Budlender, D. (2003), The debate about household headship, *Social Dynamics* **29**(2), 48–72.

Burch, T. K. and Matthews, B. J. (1987), Household formation in developed societies, *Population and Development Review* pp. 495–511.

Casale, D., Mosomi, J. and Posel, D. (forthcoming), Gender and the economy in South Africa, *The Oxford Handbook of the South African Economy* .

Casale, D. and Posel, D. (2002), The continued feminisation of the labour force in South Africa: an analysis of recent data and trends, *South African Journal of Economics* **70**(1), 156–184.

Chiappori, P.-A. et al. (1988), Rational household labor supply, *Econometrica* **56**(1), 63–90.

Chipkin, I. and Leatt, A. (2011), Religion and revival in post-apartheid South Africa, *Focus* **62**, 39–46.

Chiuri, M. C. and Del Boca, D. (2010), Home-leaving decisions of daughters and sons, *Review of Economics of the Household* **8**(3), 393–408.

Ciganda, D. and Gagnon, A. (2010), You can't go home again: independent living in Uruguay in the context of delayed transitions to adulthood, *Revista Latinoamericana de Población* **4**(6), 103–128.

Claassens, A. and Ngubane, S. (2008), Women, land and power: the impact of the Communal Land Rights Act, *in* A. Claassens and B. Cousins, eds, 'Land, Power and Custom: Controversies Generated by South Africa's Communal Land Rights Act', Cape Town University Press and Ohio University Press, pp. 154–183.

Clark, S. J., Collinson, M. A., Kahn, K., Drullinger, K. and Tollman, S. M. (2007), Returning home to die: circular labour migration and mortality in South Africa, *Scandinavian Journal of Public Health* **35**(69_suppl), 35–44.

Collinson, M. A. (2010), Striving against adversity: the dynamics of migration, health and poverty in rural South Africa, *Global Health Action* **3**(1), 5080.

Comaroff, J. and Comaroff, J. (2003), Reflections on liberalism, policulturalism, and ID-ology: citizenship and difference in South Africa, *Social Identities* **9**(4), 445–473.

Craigie, T.-A., Myers, S. L. and Darity, W. A. (2018), Racial differences in the effect of marriageable males on female family headship, *Journal of Demographic Economics* **84**(3), 231–256.

DataFirst (2015), South Africa - General Household Survey 2005 - Metadata, Technical report, University of Cape Town, Cape Town.

DataFirst (2018), Projected Population Distribution: 1990-2001, Mimeograph, University of Cape Town, Cape Town.

Dawkins, C. J. (2006), Are social networks the ties that bind families to neighborhoods?, *Housing Studies* **21**(6), 867–881.

de Lannoy, A. and Mudiriza, G. (2019), A profile of young NEETs: Unpacking the heterogeneous nature of young people not in employment, education or training in South Africa, Southern African Labour and Development Research Unit (SALDRU) Working Paper 249, University of Cape Town.

de Oliveira, O. and Salas, M. M. (2008), Social inequalities and transition to adulthood in contemporary Mexico, *Papeles de Poblacion* **14**(57), 117–152.

Deaton, A. (1997), *The analysis of household surveys: A microeconometric approach to development policy*, The Johns Hopkins University Press, United States.

Deville, J.-C. (2000), Generalized calibration and application to weighting for non-response, *in* 'COMPSTAT', Springer, pp. 65–76.

Deville, J.-C. and Särndal, C.-E. (1992), Calibration estimators in survey sampling, *Journal of the American Statistical Association* **87**(418), 376–382.

Dorrington, R., Moultrie, T. A. and Timæus, I. (2004), *Estimation of mortality using the South African Census 2001 data*, Centre for Actuarial Research, University of Cape Town: Cape Town.

DPME (2014), Twenty Year Review: South Africa 1994-2014, Background Paper: Income, Inequality and Poverty, The Presidency of South Africa, Department of Monitoring and Evaluation.

DPRU (2018), An overview of the South African labour market for the year ending 2017 quarter 4, Fact sheet 23, Development Policy Research Unit, School of Economics, University of Cape Town.

DSD (2012), The South African Child Support Grant Impact Assessment: Evidence from a survey of children, adolescents and their households, Study commissioned by, Department of Social Development (DSD); South African Social Security Agency (SASSA); and United Nations International Children's Emergency Fund (UNICEF). Pretoria: UNICEF South Africa.

Duflo, E. (2000), Child health and household resources in South Africa: evidence from the old age pension program, *American Economic Review* **90**(2), 393–398.

Ebrahim, A., Woolard, I. and Leibbrandt, M. (2013), Unemployment and household formation, SALDRU Working Paper 126, Southern Africa Labour and Development Research Unit (SALDRU); University of Cape Town, South Africa.

Edmonds, E. V., Mammen, K. and Miller, D. L. (2005), Rearranging the family? Income support and elderly living arrangements in a low-income country, *Journal of Human Resources* **40**(1), 186–207.

Ermisch, J. (1981), An economic theory of household formation: Theory and evidence from the General Household Survey, *Scottish Journal of Political Economy* **28**(1), 1–19.

Ermisch, J. (1999), Prices, parents, and young people's household formation, *Journal of Urban Economics* **45**(1), 47–71.

Ermisch, J. and Di Salvo, P. (1996), Surprises and housing tenure decisions in Great Britain, *Journal of Housing Economics* **5**(3), 247–273.

Esping-Andersen, G. and Billari, F. C. (2015), Re-theorizing family demographics, *Population and Development Review* **41**(1), 1–31.

Esteve, A., Reher, D. S., Treviño, R., Zueras, P. and Turu, A. (2020), Living alone over the life course: Cross-national variations on an emerging issue, *Population and Development Review* **46**(1), 169–189.

Evans, A. (2013), Generational change in leaving the parental home, *in* 'Negotiating the Life Course: Stability and Change in Life Pathways', Springer Netherlands, Dordrecht, pp. 53–67.

Ezra, M. (2000), Leaving-home of young adults under conditions of ecological stress in the drought prone communities of northern Ethiopia, *Genus* pp. 121–144.

Fafchamps, M. and Quisumbing, A. R. (2007), Household formation and marriage markets in rural areas, *in* 'Handbook of development economics', Vol. 4, Elsevier, chapter 51, pp. 3187–3247.

Fairlie, R. W. (2005), An extension of the Blinder-Oaxaca decomposition technique to logit and probit models, *Journal of Economic and Social Measurement* **30**(4), 305–316.

Finn, A., Leibbrandt, M. and Woolard, I. (2013), What happened to multidimensional poverty in South Africa between 1993 and 2010?, SALDRU Working Paper No. 99, Southern Africa Labour and Development Research Unit (SALDRU), University of Cape Town, South Africa.

Folbre, N. (1995), "Holding hands at midnight": The paradox of caring labor, *Feminist Economics* **1**(1), 73–92.

Fortin, N. (2006), Greed, altruism, and the gender wage gap, *Unpublished manuscript, Department of Economics, University of British Columbia* .

Fortin, N., Lemieux, T. and Firpo, S. (2011), Decomposition methods in economics, *in* 'Handbook of Labor Economics', Vol. 4, Elsevier, chapter 1, pp. 1–102.

Gardeazabal, J. and Ugidos, A. (2004), More on identification in detailed wage decompositions, *Review of Economics and Statistics* **86**(4), 1034–1036.

Garenne, M. (2016), A century of nuptiality decline in South Africa: A longitudinal analysis of census data, *African Population Studies* **30**(2).

Golan, A., Judge, G. and Miller, D. (1997), The maximum entropy approach to estimation and inference, *in* 'Applying Maximum Entropy to Econometric Problems', Emerald Group Publishing Limited, pp. 3–24.

Goldberg, R. E. (2013), Family instability and pathways to adulthood in Cape Town, South Africa, *Population and Development Review* **39**(2), 231–256.

Goldscheider, F. (1997), Recent changes in us young adult living arrangements in comparative perspective, *Journal of Family Issues* **18**(6), 708–724.

Goldscheider, F. and DaVanzo, J. (1989), Pathways to independent living in early adulthood: Marriage, semiautonomy, and premarital residential independence, *Demography* **26**(4), 597–614.

Goode, W. J. (1963), *World revolution and family patterns.*, Free Press Glencoe.

Gouskova, E., Chiteji, N. and Stafford, F. (2010), Estimating the intergenerational persistence of lifetime earnings with life course matching: Evidence from the PSID, *Labour Economics* **17**(3), 592–597.

Grant, M. J. and Furstenberg, F. F. (2007), Changes in the transition to adulthood in less developed countries, *European Journal of Population* **23**(3/4), 415–428.

Greenwood, J., Seshadri, A. and Yorukoglu, M. (2005), Engines of liberation, *The Review of Economic Studies* **72**(1), 109–133.

Grossbard, S. (2006), The New Home Economics at Columbia and Chicago, *in* 'Jacob Mincer: A Pioneer of Modern Labor Economics', Springer, pp. 37–49.

Hall, K. and Posel, D. (2019), Fragmenting the Family? The complexity of household migration strategies in post-apartheid South Africa, *IZA Journal of Development and Migration* **10**(2).

Hanna, K. L. and Collins, P. F. (2015), Relationship between living alone and food and nutrient intake, *Nutrition Reviews* **73**(9), 594–611.

Harris, T., Collinson, M. and Wittenberg, M. (2017), Aiming for a moving target: the dynamics of household electricity connections in a developing context, *World Development* **97**, 14–26.

Hatch, M. and Posel, D. (2018), Who cares for children? A quantitative study of childcare in South Africa, *Development Southern Africa* **35**(2), 267–282.

Heckman, J. J. (2015), Introduction to: A theory of the allocation of time by Gary Becker, *The Economic Journal* **125**(583), 403–409.

Helman, R. and Ratele, K. (2016), Everyday (in)equality at home: complex constructions of gender in South African families, *Global Health Action* **9**(1), 31122.

Hendrix, L. (1975), Nuclear family universals: Fact and faith in the acceptance of an idea, *Journal of Comparative Family Studies* **6**(2), 125–138.

Hofferth, S. L., Boisjoly, J. and Duncan, G. J. (1998), Parents' extrafamilial resources and children's school attainment, *Sociology of Education* pp. 246–268.

Hoge, D. R., Petrillo, G. H. and Smith, E. I. (1982), Transmission of religious and social values from parents to teenage children, *Journal of Marriage and the Family* pp. 569–580.

Holt-Lunstad, J., Smith, T. B. and Layton, J. B. (2010), Social relationships and mortality risk: a meta-analytic review, *PLoS Medicine* **7**(7).

Honwana, A. (2014), 'Waithood': Youth Transitions and Social Change, *in* 'Development and Equity', Brill, chapter 2, pp. 28–40.

Honwana, A. M. (2012), *The time of youth: Work, social change, and politics in Africa*, Kumarian Press Sterling.

Horrace, W. C. and Oaxaca, R. L. (2006), Results on the bias and inconsistency of ordinary least squares for the linear probability model, *Economics Letters* **90**(3), 321–327.

Hosegood, V., McGrath, N. and Moultrie, T. (2009), Dispensing with marriage: Marital and partnership trends in rural KwaZulu-Natal, South Africa 2000-2006, *Demographic Research* **20**, 279.

Hunter, M. (2002), The materiality of everyday sex: thinking beyond 'prostitution', *African Studies* **61**(1), 99–120.

Hunter, M. (2010), *Love in the time of AIDS: inequality, gender, and rights in South Africa*, Indiana University Press.

Hunter, M. and Posel, D. (2012), Here to work: the socioeconomic characteristics of informal dwellers in post-apartheid south africa, *Environment and Urbanization* **24**(1), 285–304.

Jamieson, L., Wasoff, F. and Simpson, R. (2009), Solo-living, demographic and family change: The need to know more about men, *Sociological Research Online* **14**(2), 20–35.

Jann, B. (2008), The Blinder–Oaxaca decomposition for linear regression models, *The Stata Journal* **8**(4), 453–479.

Jann, B. (2018), Decomposition methods in the social sciences [slides], *in* 'Workshop at the Bamberg Graduate School of Social Sciences'. June 7-8.

Jhamba, T. and Mmatli, K. (2015), Household and family dynamics in south africa, *in* 'Social demography of South Africa: Advances and emerging issues', chapter 9.

Johnson, R. W. and DaVanzo, J. (1998), Economic and cultural influences on the decision to leave home in peninsular Malaysia, *Demography* **35**(1), 97–114.

Jones, F. L. and Kelley, J. (1984), Decomposing differences between groups: A cautionary note on measuring discrimination, *Sociological Methods & Research* **12**(3), 323–343.

Kalleberg, A. L. and Vallas, S. P. (2018), Probing precarious work: Theory, research, and politics, *Research in the Sociology of Work* **31**(1), 1–30.

Karim, Q. A. and Baxter, C. (2016), The dual burden of gender-based violence and HIV in adolescent girls and young women in South Africa, *South African Medical Journal* **106**(12), 1151–1153.

Karimakwenda, N. (2020), The language of erasure: neutralization techniques in contemporary South African marital rape judgments, *Violence against Women* .

Kato, T. A., Shinfuku, N., Sartorius, N. and Kanba, S. (2017), Loneliness and single-person households: Issues of kodokushi and hikikomori in Japan, *in* 'Mental health and illness worldwide: Mental health and illness in the city', Springer Science + Business Media, pp. 205–219.

Keller, S. (2004), Household formation, poverty and unemployment—the case of rural households in South Africa, *South African Journal of Economics* **72**(3), 437–483.

Kerr, A., Lam, D. and Wittenberg, M. (2019), Post-Apartheid Labour Market Series: 1993-2019 [dataset], University of Cape Town: DataFirst [producer and distributor]. Version 3.3.

Kerr, A. and Wittenberg, M. (2015), Sampling methodology and fieldwork changes in the October Household Surveys and Labour Force Surveys, *Development Southern Africa* **32**(5), 603–612.

King, G. and Zeng, L. (2001), Logistic regression in rare events data, *Political Analysis* **9**(2), 137–163.

Klasen, S. and Woolard, I. (2009), Surviving unemployment without state support: unemployment and household formation in South Africa, *Journal of African Economies* **18**(1), 1–51.

Klinenberg, E. (2013), *Going solo: The extraordinary rise and surprising appeal of living alone*, Penguin.

Klinenberg, E. (2016), Social isolation, loneliness, and living alone: identifying the risks for public health, *American Journal of Public Health* **106**(5), 786.

Lam, D. and Schoeni, R. F. (1993), Effects of family background on earnings and returns to schooling: evidence from Brazil, *Journal of Political Economy* **101**(4), 710–740.

Lanjouw, P. and Ravallion, M. (1995), Poverty and household size, *The Economic Journal* **105**(433), 1415–1434.

Lavallée, P. and Beaumont, J.-F. (2015), Why we should put some weight on weights, *Survey Methods: Insights from the Field* . Special Issue 'Weighting: Practical Issues and 'How to' Approach'. Invited Article.

Leibbrandt, M., Finn, A. and Oosthuizen, M. (2016), Poverty, inequality, and prices in post-apartheid South Africa, *in* 'Growth and Poverty in sub-Saharan Africa', Vol. 1, Oxford University Press, p. 393.

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, No. 101.

Leitgöb, H. (2020), Analysis of rare events, *in* 'SAGE Research Methods Foundations'.

Lemon, A. (1991), *Homes apart: South Africa's segregated cities*, Indiana University Press.

Lesthaeghe, R. (2010), The unfolding story of the second demographic transition, *Population and Development Review* **36**(2), 211–251.

Lesthaeghe, R. (2020), The second demographic transition: Cohabitation, *in* 'Cross-Cultural Family Research and Practice', Elsevier, pp. 103–141.

Lichter, D. T. and Qian, Z. (2019), The study of assortative mating: Theory, data, and analysis, *in* 'Analytical Family Demography', Springer, pp. 303–337.

Lundberg, S. (2001), Gender and household decision-making, *in* 'Frontiers in the Economics of Gender', Routledge, pp. 132–150.

Lundberg, S. and Pollak, R. A. (1993), Separate spheres bargaining and the marriage market, *Journal of Political Economy* **101**(6), 988–1010.

Mabin, A. (2003), Dispossession, exploitation and struggle: an historical overview of south african urbanization, *in* 'The apartheid city and beyond', Routledge, pp. 24–36.

Machemedze, T., Kerr, A. and Wittenberg, M. (2007), Recalibrating the OHSs to adjust for sampling changes, DataFirst Technical Paper No. 28, DataFirst, University of Cape Town, South Africa.

Madhavan, S., Harrison, A. and Sennott, C. (2013), Management of non-marital fertility in two south african communities, *Culture, health & sexuality* **15**(5), 614–628.

Madhavan, S., Townsend, N. W. and Garey, A. I. (2008), 'Absent breadwinners': Father–child connections and paternal support in rural South Africa, *Journal of Southern African Studies* **34**(3), 647–663.

Maisiri, O. (2016), Male headship and female submission in marriage among african christians: a pastoral perspective, Master's thesis, University of Pretoria.

Makusha, T., Richter, L., Naicker, S., Swartz, S. and Desmon, C. (2019), Regarding Baba: perceptions about fathers and fatherhood in South Africa, *in* 'Famiy Matters: family cohesion, values and strengthening to promote well-being', HSRC Press, Cape Town, chapter 11.

Malinowski, B. (1930), Parenthood, the basis of social structure, *in* 'The New Generation', New York: Macaulay.

Marais, L. and Venter, A. (2006), Hating the compound, but... mineworker housing needs in post-apartheid South Africa, *Africa Insight* **36**(1), 53–62.

McCue, D., Masnick, G. and Herbert, C. (2015), Assessing households and household growth estimates with Census Bureau surveys, Working Paper W15-5, Cambridge, MA: Joint Center for Housing Studies of Harvard University.

McDonald, P. (1992), Convergence or compromise in historical family change?, *Family Systems and Cultural Change* pp. 15–30.

McElroy, M. B. (1985), The joint determination of household membership and market work: The case of young men, *Journal of Labor Economics* **3**(3), 293–316.

McElroy, M. B. and Horney, M. J. (1981), Nash-bargained household decisions: Toward a generalization of the theory of demand, *International Economic Review* pp. 333–349.

Mhlongo, V. and Daniels, R. C. (2013), Food expenditure patterns in South Africa: Evidence from the NIDS, SALDRU Working Paper No. 123, Southern Africa Labour and Development Research Unit (SALDRU), University of Cape Town.

Mincer, J. (1962), Labor force participation of married women: A study of labor supply, *in* 'Aspects of Labor Economics', Princeton University Press, pp. 63–105.

Mohlabane, N., Gumede, N. and Mokomane, Z. (2019), Attitudes towards marriage in post-apartheid South Africa, *in* 'Famiy Matters: family cohesion, values and strengthening to promote well-being', HSRC Press, Cape Town, chapter 9.

Molgat, M. (2002), Leaving home in Quebec: theoretical and social implications of (im)mobility among youth, *Journal of Youth Studies* **5**(2), 135–152.

Mosomi, J. (2019), An empirical analysis of trends in female labour force participation and the gender wage gap in South Africa, *Agenda* **33**(4), 29–43.

Murdock, G. P. (1949), *Social Structure*, Macmillan.

Musinguzi, L. K. (2016), The role of social networks in savings groups: insights from village savings and loan associations in Luwero, Uganda, *Community Development Journal* **51**(4), 499–516.

Mutanda, N. and Odimegwu, C. O. (2019), Solitary living in South Africa: what is driving the pattern and change?, *Journal of Population Research* **36**(2), 137–158.

National Treasury (2004), Intergovernmental fiscal review 2000/01-2006/07: Chapter 9 - housing, Technical report, Government of South Africa, Pretoria.

Nattrass, N. and Seekings, J. (2010), The economy and poverty in the twentieth century in South Africa, CSSR Working Paper No. 276, Centre for Social Science Research (CSSR), University of Cape Town.

Neumark, D. (1988), Employers' discriminatory behavior and the estimation of wage discrimination, *Journal of Human Resources* pp. 279–295.

Nolde, J. (1991), South african women under apartheid: Employment rights with particular focus on domestic service and forms of resistance to promote change, *Third World Legal Stud.* p. 203.

Nyasani, E., Sterberg, E. and Smith, H. (2009), Fostering children affected by AIDS in Richards Bay, South Africa: A qualitative study of grandparents' experiences, *African Journal of AIDS Research* **8**(2), 181–192.

Oaxaca, R. L. (1973), Male-female wage differentials in urban labor markets, *International Economic Review* pp. 693–709.

Oaxaca, R. L. and Ransom, M. R. (1994), On discrimination and the decomposition of wage differentials, *Journal of Econometrics* **61**(1), 5–21.

Oaxaca, R. L. and Ransom, M. R. (1999), Identification in detailed wage decompositions, *Review of Economics and Statistics* **81**(1), 154–157.

OECD (2011), Families are changing, *in* 'Doing better for families', chapter 1.

Ortiz-Ospina, E. (2019), The rise of living alone: how one-person households are becoming increasingly common around the world, *Our World in Data* .
URL: *https://ourworldindata.org/living-alone*

Pantell, M., Rehkopf, D., Jutte, D., Syme, S. L., Balmes, J. and Adler, N. (2013), Social isolation: a predictor of mortality comparable to traditional clinical risk factors, *American Journal of Public Health* **103**(11), 2056–2062.

Parnell, S. and Crankshaw, O. (2013), The politics of 'race'and the transformation of the post-apartheid space economy, *Journal of Housing and the built Environment* **28**(4), 589–603.

Peichl, A., Pestel, N. and Schneider, H. (2012), Does size matter? The impact of changes in household structure on income distribution in Germany, *Review of Income and Wealth* **58**(1), 118–141.

Pérez Amador, J. (2006), El inicio de la vida laboral como detonador de la independencia residencial de los jóvenes en méxico, *Estudios demográficos y urbanos* **21**(1), 7–47.

Pietromonaco, P. R. and Collins, N. L. (2017), Interpersonal mechanisms linking close relationships to health., *American Psychologist* **72**(6), 531.

Pirouz, F. (2005), Have labour market outcomes affected household structure in South Africa? A descriptive analysis of households, Working paper no. 05/100, Development Policy Research Unit, University of Cape Town.

Pleschberger, S., Reitinger, E., Trukeschitz, B. and Wosko, P. (2019), Older people living alone (OPLA)–non-kin-carers' support towards the end of life: qualitative longitudinal study protocol, *BMC geriatrics* **19**(1), 219.

Pollak, R. A. (1985), A transaction cost approach to families and households, *Journal of economic Literature* **23**(2), 581–608.

Portacolone, E., Rubinstein, R. L., Covinsky, K. E., Halpern, J. and Johnson, J. K. (2019), The precarity of older adults living alone with cognitive impairment, *The Gerontologist* **59**(2), 271–280.

Posel, D. (2001), Who are the heads of household, what do they do, and is the concept of headship useful? An analysis of headship in South Africa, *Development Southern Africa* **18**(5), 651–670.

Posel, D. (2003), The collection of national household survey data in South Africa (1993-2001): Rendering labour migration invisible, *Development Southern Africa* **20**(3), 361–368.

Posel, D. (2010), Households and labour migration in post-apartheid South Africa, *Studies in Economics and Econometrics* **34**(3), 129–141.

Posel, D. (2020), Measuring labour migration after apartheid: Patterns and trends, *in* 'Migrant labour after apartheid: the inside story.', HSRC Press, Cape Town, chapter 2.

Posel, D. and Casale, D. (2003), What has been happening to internal labour migration in South Africa, 1993–1999?, *South African Journal of Economics* **71**(3), 455–479.

Posel, D. and Casale, D. (2013), The relationship between sex ratios and marriage rates in South Africa, *Applied Economics* **45**(5), 663–676.

Posel, D. and Casale, D. (2020), Who moves during times of crisis? Mobility, living arrangements and COVID-19 in South Africa, National Income Dynamics Study (NIDS) - Coronavirus Rapid Mobile Survey (CRAM) Working Paper 8.

Posel, D. and Hall, K. (forthcoming), Families and household formation in south africa, *The Oxford Handbook of the South African Economy* .

Posel, D. and Marx, C. (2013), Circular migration: a view from destination households in two urban informal settlements in South Africa, *The Journal of Development Studies* **49**(6), 819–831.

Posel, D. and Rogan, M. (2012), Gendered trends in poverty in the post-apartheid period, 1997–2006, *Development Southern Africa* **29**(1), 97–113.

Posel, D. and Rogan, M. (2016), Measured as poor versus feeling poor: Comparing money-metric and subjective poverty rates in South Africa, *Journal of Human Development and Capabilities* **17**(1), 55–73.

Posel, D. and Rudwick, S. (2013), Changing patterns of marriage and cohabitation in South Africa, *Acta Juridica* **2013**(1), 169–180.

Posel, D. and Rudwick, S. (2014), Marriage and Bridewealth (Ilobolo) in contemporary Zulu society, *African Studies Review* **57**(2), 51–72.

Powers, D. A., Yoshioka, H. and Yun, M.-S. (2011), mvdcmp: Multivariate decomposition for nonlinear response models, *The Stata Journal* **11**(4), 556–576.

President Motlanthe (2009), State of the Nation Address. Speech by the President of South Africa to Parliament, 6 February.
URL: *https://www.sahistory.org.za/archive/2009-president-motlanthe-state-nation-address-6-february-2009-national-elections*

Presser, H. B. (1998), Decapitating the US Census Bureau's "head of household": Feminist mobilization in the 1970s, *Feminist Economics* **4**(3), 145–158.

Rabe, M. (2006), Being a father in a man's world: the experiences of goldmine workers, *in* 'Baba: men and fatherhood in South Africa', HSRC Press Cape Town, pp. 250–264.

Ranchhod, V. (2017), Household responses to the cessation of grant income: The case of south africa's old age pension, Working Paper 40, Research Project on Employment, Income Distribution and Inclusive Growth (REDI3x3), University of Cape Town, South Africa.

Raymo, J. (2015), Living alone in Japan: Relationships with happiness and health, *Demographic Research* **32**, 1267.

Raymo, J. and Ono, H. (2007), Coresidence with parents, women's economic resources, and the transition to marriage in Japan, *Journal of Family Issues* **28**(5), 653–681.

Reher, D. (1998), Family ties in Western Europe: persistent contrasts, *Population and Development Review* pp. 203–234.

Reher, D. and Requena, M. (2018), Living alone in later life: A global perspective, *Population and Development Review* **44**(3), 427–454.

Rindfuss, R. R. (1991), The young adult years: Diversity, structural change, and fertility, *Demography* **28**(4), 493–512.

Rogan, M. (2013), Poverty and headship in post-apartheid South Africa, 1997–2006, *Social Indicators Research* **113**(1), 491–511.

Rogan, M. (2016), Qualitative perceptions of the meaning of "headship" and female-headed households in post-apartheid South Africa, *Social Dynamics* **42**(1), 175–195.

Rommelspacher, A. (2020), "Too delicate to work"? Breadwinner, women and work in Cape Town 1938-1939. Mimeograph. Stellenbosch University.

Rosenzweig, M. R. and Wolpin, K. I. (1993), Intergenerational support and the life-cycle incomes of young men and their parents: Human capital investments, coresidence, and intergenerational financial transfers, *Journal of Labor Economics* **11**(1, Part 1), 84–112.

Rosenzweig, M. R. and Wolpin, K. I. (1994), Parental and public transfers to young women and their children, *The American Economic Review* pp. 1195–1212.

Roux, D. J. and Geyer, H. S. (2017), Demographic transitions in South African cities: An analysis of household structures in the City of Tshwane, *Regional Science Policy & Practice* **9**(3), 165–181.

Russell, M. (1994), Do Blacks live in nuclear family households? An appraisal of Steyn's work on urban family structure in South Africa, *South African Sociological Review* pp. 56–67.

Russell, M. (1998), Black urban households in South Africa, *African Sociological Review* **2**(1), 174–180.

Russell, M. (2003), Are urban black families nuclear? A comparative study of black and white South African family norms, *Social Dynamics* **29**(2), 153–176.

Russell, M. (2004), Understanding black households in southern africa: The african kinship and western nuclear family systems, CSSR Working Paper No. 67, Centre for Social Science Research (CSSR), Social Surveys Unit.

Salcedo, A., Schoellman, T. and Tertilt, M. (2012), Families as roommates: Changes in us household size from 1850 to 2000, *Quantitative Economics* **3**(1), 133–175.

SALDRU (2016), National Income Dynamics Study - Waves 1-4 (2000/2010-11/2012/2014-5) [dataset], Pretoria: SA Presidency [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer]. Cape Town: DataFirst [distributor]. Wave 1 (Version 7), Wave 2 (Version 4), Wave 3 (Version 3), Wave 4 (Version 1.1).

Sandefur, G. D., Meier, A. M. and Campbell, M. E. (2006), Family resources, social capital, and college attendance, *Social Science Research* **35**(2), 525–553.

Särndal, C.-E. (2007), The calibration approach in survey theory and practice, *Survey Methodology* **33**(2), 99–119.

Schöer, V., Rankin, N. and Roberts, G. (2014), Accessing the first job in a slack labour market: Job matching in South Africa, *Journal of International Development* **26**(1), 1–22.

Schwartz, C. R. (2013), Trends and variation in assortative mating: Causes and consequences, *Annual Review of Sociology* **39**, 451–470.

Seekings, J. (2008), Beyond 'fluidity': Kinship and households as social projects, CSSR Working Paper No. 237, Centre for Social Science Research (CSSR), Social Surveys Unit, University of Cape Town, South Africa.

Settersten, R. A. (2003), Age structuring and the rhythm of the life course, *in* 'Handbook of the life course', Springer, pp. 81–98.

Shefer, T. (2010), Narrating gender and sex in and through apartheid divides, *South African Journal of Psychology* **40**(4), 382–395.

Shih, V. (2004), Factions matter: personal networks and the distribution of bank loans in china, *Journal of Contemporary China* **13**(38), 3–19.

Simkins, C. (1986), Household Composition and Structure in South Africa, *in* 'Growing up in a Divided Society', Johannesburg: Raven Press.

Simkins, C. (2017*a*), Staying at home and leaving I - measurement, HSF Brief, Helen Suzman Foundation (HSF).
URL: *http://hsf.org.za/resource-centre/hsf-briefs/staying-at-home-and-leaving-i-measurement*

Simkins, C. (2017*b*), Staying at home and leaving II - determinants, HSF Brief, Helen Suzman Foundation (HSF).
URL: *http://hsf.org.za/resource-centre/hsf-briefs/staying-at-home-and-leaving-ii-determinants*

Siqwana-Ndulo, N. (1998), Rural African family structure in the Eastern Cape Province, South Africa, *Journal of Comparative Family Studies* **29**(2), 407–417.

Sironi, M., Barban, N., Pesando, L. M. and Furstenberg, F. F. (2020), A sequence-analysis approach to the study of the transition to adulthood in low-and middle-income countries, Population center working papers (psc/parc), Population Studies Centre at University of Pennsylvania.

Smit, R. (2001), The impact of labor migration on African families in South Africa: Yesterday and today, *Journal of Comparative Family Studies* **32**(4), 533–548.

Smith, D. M. (2003), *The apartheid city and beyond: Urbanization and social change in South Africa*, Routledge.

Smith, K. J. and Victor, C. (2019), Typologies of loneliness, living alone and social isolation, and their associations with physical and mental health, *Ageing & Society* **39**(8), 1709–1730.

Smith, T. M. (1991), Post-stratification, *Journal of the Royal Statistical Society: Series D (The Statistician)* **40**(3), 315–323.

Snell, K. (2017), The rise of living alone and loneliness in history, *Social History* **42**(1), 2–28.

Spiegel, A., Watson, V. and Wilkinson, P. (1996), Domestic diversity and fluidity among some African households in Greater Cape Town, *Social Dynamics* **22**(1), 7–30.

StatsSA (1996), South African Census 1996, 10% Sample: Households Metadata, Technical report, Statistics South Africa (StatsSA). DataFirst.

StatsSA (1997), October Household Survey 1997: Metadata, Technical report, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (1998), October Household Survey 1998: Metadata, Technical report, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (1999), October Household Survey 1999: Metadata, Technical report, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (2001*a*), Census 2001: Primary tables South Africa: Census '96 and 2001 compared, Report No. 03-02-04, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (2001*b*), Investigation into appropriate definitions of urban and rural areas for South Africa, Discussion document Report no. 03-02-20, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (2007*a*), Community Survey 2007: Basic Results, Pamphlet, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (2007*b*), Community Survey 2007: Unit Records Metadata, Technical report, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (2008*a*), General Household Survey 2008 - Statistical Release P0318, Technical report, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (2008*b*), Measuring poverty in South Africa: Methodological report on the development of the poverty lines for statistical reporting, Technical Report No. D0300, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (2008*c*), Rent in the South African CPI: concepts and trends, Technical report, Statistics South Africa (StatsSA), Government of South Africa, Pretoria.

StatsSA (2010-2013), October Household Survey: 1995-1999 [datasets], Pretoria: Statistics South Africa (StatsSA) [producer]. University of Cape Town: DataFirst [distributor]. Version 1.1 [1996-1999]; Version 1.2 [1995].

StatsSA (2010*a*), Community Survey 2007 [dataset]., Pretoria: Statistics South Africa (StatsSA) [producer], 2008. Cape Town: DataFirst [distributor], 2010. DOI: https://doi.org/10.25828/0nqv-ns26. Version 1.2.

StatsSA (2010*b*), General Household Survey 2009 - Statistical Release P0318, Technical report, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (2010*c*), Reweighting of the GHS 2002–2008 data series, Technical report, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (2011), South African Census 2001, 10% sample [dataset]., Pretoria: Statistics South Africa (StatsSA) [producer], 2003. Cape Town: DataFirst [distributor], 2011. Version 1.1.

StatsSA (2011-2018*b*), General Household Survey: 2002-2015 [datasets], Pretoria: Statistics South Africa (StatsSA) [producer]. University of Cape Town: DataFirst [distributor]. Version 1 [2013]; Version 1.1 [2011, 2014]; Version 1.2 [2002, 2015]; Version 1.3 [2003-2009]; Version 2.1 [2010]; Version 2 [2012].

StatsSA (2012*a*), Census 2011: 10% Sample Metadata, Technical report, Statistics South Africa (StatsSA). DataFirst.

StatsSA (2012*b*), Census 2011: Post-enumeration survey: Results and methodology, Report No. 03-01-46, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (2012*c*), Census 2011: Statistical release (revised), Statistical release No. P0301.4, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa.

StatsSA (2015*a*), South African Census 1996, 10% sample [dataset]., Pretoria: Statistics South Africa (StatsSA) [producer], 1998. Cape Town: DataFirst [distributor], 2015. DOI: https://doi.org/10.25828/7yrd-n169. Version 1.2.

StatsSA (2015*b*), South African Census 2011, 10% sample [dataset]., Pretoria: Statistics South Africa (StatsSA) [producer], 2015. Cape Town: DataFirst [distributor], 2015. DOI: https://doi.org/10.25828/vjy1-tz66. Version 2.

StatsSA (2018*a*), Country projection by population group, sex and age (2002-2017) [Excel file], 'Additional download' related to publication P0301-Mid-year population estimates, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa. Retrieved from http://www.statssa.gov.za/?page$_id = 1854PPN = P0302$.

StatsSA (2018*c*), Provincial projection by sex and age (2002-2017) [Excel file], 'Additional download' related to publication P0301-Mid-year population estimates, Statistics South Africa (StatsSA), Government of South Africa, Pretoria, South Africa. Retrieved from http://www.statssa.gov.za/?page$_id = 1854PPN = P0302$.

StatsSA (2019), Inequality trends in South Africa: a multidimensional diagnostic of inequality, Report No. 03-10-19, Statistics South Africa (StatsSA), Government of South Africa; Southern African Labour Development Research Unit (SALDRU); Agence Francaise de Developpement (AFD), Pretoria, South Africa.

Steyn, A. (1995), Urban household structures in the Republic of South Africa, *Family Formations and Dissolution: perspectives from East and West* **36**, 169–204.

Sung, H. (2020), Identification on the differentiating characteristics of determinant factors on commuting mode choice for the single-person household compared to the multi-person household, *LHI Journal of Land, Housing, and Urban Affairs* **11**(2), 1–14.

Suzuki, T. (2001), Leaving the parental household in contemporary Japan, *Review of Population and Social Policy* **10**, 23–35.

Thurman, S. (1997), Umzamo: Improving hostel dwellers' accommodation in South Africa, *Environment and Urbanization* **9**(2), 43–62.

Timol, F., Lynch, I. and Morison, T. (2019), Is a woman's place still in the home? Gender-role attitudes and women's position in the South African labour market, *in* 'Famiy Matters: family cohesion, values and strengthening to promote well-being', HSRC Press, Cape Town, chapter 10.

Timoneda, J. C. (2020), Estimating group fixed effects in panel data with a binary dependent variable: How the LPM outperforms logistic regression in rare events data, *Social Science Research* **93**, 102486.

Ting, K.-f. and Chiu, S. W. (2002), Leaving the parental home: Chinese culture in an urban context, *Journal of Marriage and Family* **64**(3), 614–626.

UN (2007), Enhancing urban safety and security: global report on human settlements, Technical report, United Nations (UN) Human Settlements Programme.

UN (2017), Household size and composition around the world 2017: Data booklet, Data Booklet (ST/ESA/ SER.A/405), United Nations (UN) Department of Economic and Social Affairs, Population Division, New York.

UNFPA (2019), World population dashboard, United Nations Fund for Population Activities (UNFPA) New York.
URL: *https://www.unfpa.org/data/world-population-dashboard*

van der Berg, S. (2007), Apartheid's enduring legacy: Inequalities in education, *Journal of African Economies* **16**(5), 849–880.

van der Berg, S., Siebrits, F. K. and Lekezwa, B. (2010), Efficiency and equity effects of social grants in South Africa, Stellenbosch Economic Working Papers No. 15/10, Stellenbosch University and Bureau for Economic Research.

van der Paal, B. (2014), A comparison of different methods for modelling rare events data, Master's thesis, Ghent University.

Venter, C. M. (1995), The new South African constitution: Facing the challenges of women's rights and cultural rights in post-apartheid South Africa, *Journal of Legislation* **21**, 1.

Vetten, L. (2013), Deserving and undeserving women: A case study of South African policy and legislation addressing domestic violence, Master's thesis, University of the Witwatersrand.

Wilson, F. (2001), Minerals and migrants: how the mining industry has shaped South Africa, *Daedalus* **130**(1), 99–121.

Winsborough, H. and Dickinson, P. (1971), Components of negro-white income differences, *Age* **25**(34), 35–44.

Wittenberg, M. (2008), October Household Survey 1994, Technical report, DataFirst, University of Cape Town.
URL: *https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/407/download/5258*

Wittenberg, M. (2010), An introduction to maximum entropy and minimum cross-entropy estimation using Stata, *Stata Journal* **10**(3), 315.

Wittenberg, M. (2017), Wages and wage inequality in South Africa 1994–2011: Part 2–inequality measurement and trends, *South African Journal of Economics* **85**(2), 298–318.

Wittenberg, M. and Collinson, M. A. (2007), Household transitions in rural South Africa, 1996—2003, *Scandinavian Journal of Public Health* **35**(69_suppl), 130–137.

Wittenberg, M. and Collinson, M. A. (2020), Household formation and service delivery in post-apartheid South Africa: Evidence from the Agincourt sub-district 1992–2012, *Development Southern Africa* pp. 1–19.

Wittenberg, M., Collinson, M. A. and Harris, T. (2017), Decomposing changes in household measures: Household size and services in South Africa, 1994–2012, *Demographic Research* **37**, 1297–1326.

Woolard, I. and Klasen, S. (2005), Determinants of income mobility and household poverty dynamics in South Africa, *Journal of Development Studies* **41**(5), 865–897.

Xulu, N. (2014), From hostels to CRUs: Spaces of perpetual perplexity, *South African Review of Sociology* **45**(1), 140–154.

Yi, Z., Coale, A., Choe, M. K., Zhiwu, L. and Li, L. (1994), Leaving the parental home: Census-based estimates for China, Japan, South Korea, United States, France, and Sweden, *Population Studies* **48**(1), 65–80.

Yun, M.-S. (2004), Decomposing differences in the first moment, *Economics letters* **82**(2), 275–280.

Yun, M.-S. (2005), A simple solution to the identification problem in detailed wage decompositions, *Economic inquiry* **43**(4), 766–772.

Zembe, Y. and Adjiwanou, V. (2019), Intimate partner violence and family cohesion among ever married women and men in South Africa, *in* 'Famiy Matters: family cohesion, values and strengthening to promote well-being', HSRC Press, Cape Town, chapter 6.

Ziehl, S. (2001), Documenting changing family patterns in South Africa: Are census data of any value?, *African Sociological Review* **5**(2), 36–62.

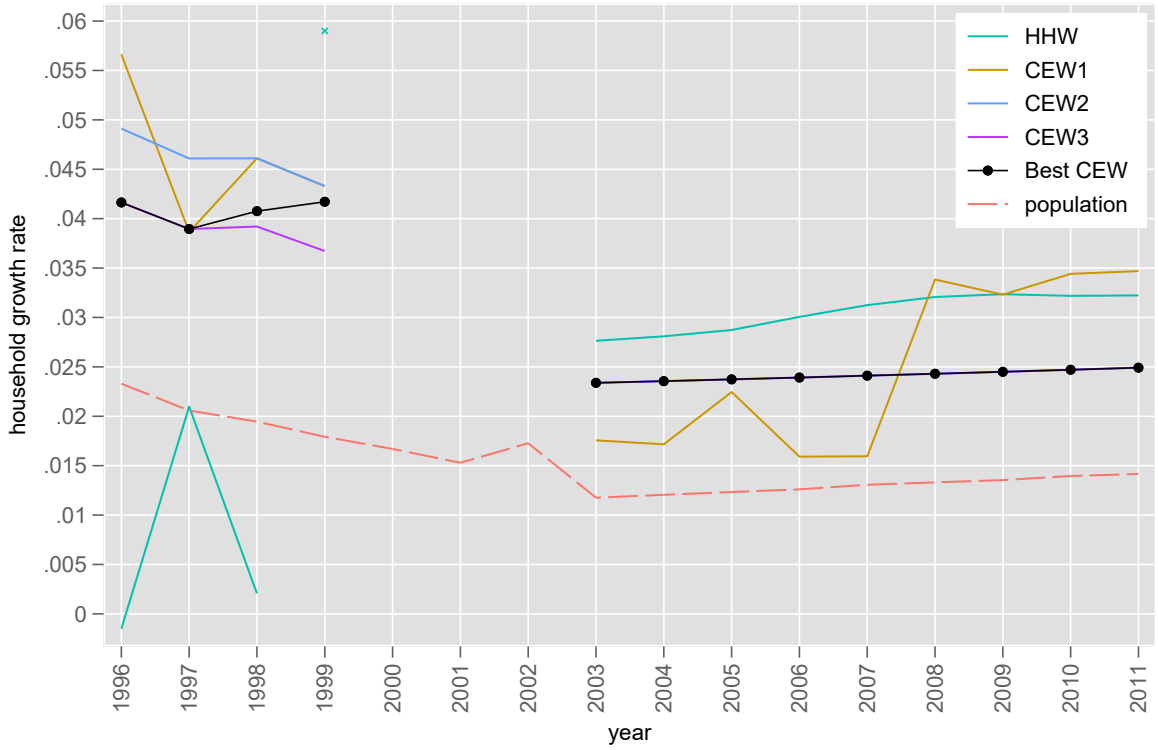
Ziehl, S. (2002), Black South Africans do live in nuclear family households—a response to Russell, *Society in Transition* **33**(1), 26–49.

Appendix A

Additional material for Chapter 5

A.1 Population and household growth rates

Figure A.1: Population and household growth rates, 1995-2011



Notes: population benchmarks from DataFirst (2018) (1995-2002) and StatsSA (2018*a*). Household growth rates are own calculations using the OHS-GHS series weighted: HHW = household weight; CEW1-Best CEW = cross-entropy weight series described in Table 4.1. x = the HHW growth rate estimate is cut off in 1999 because it is out of the plot's range at 0.16

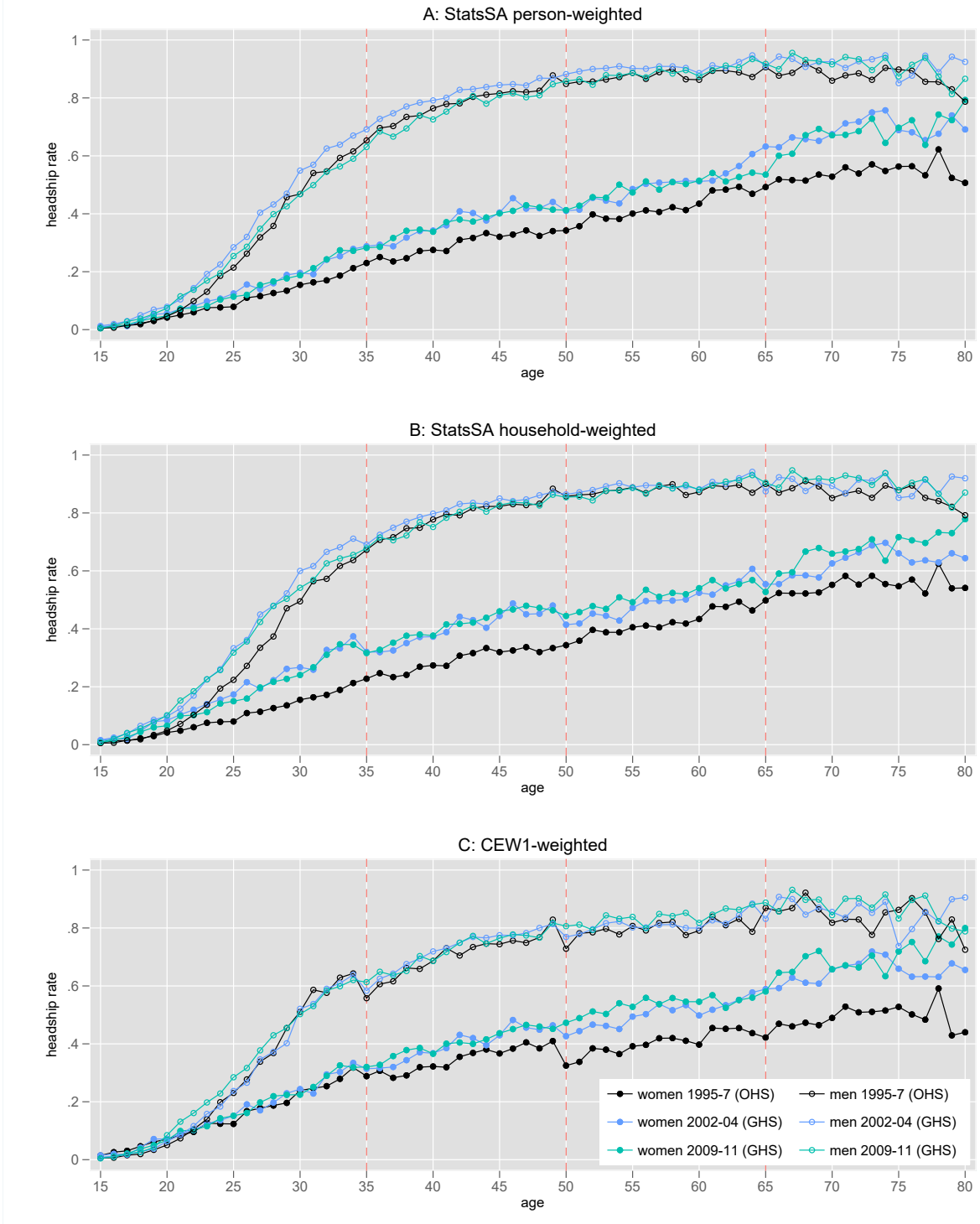
A.2 Impact of different weights on the age-profile of headship and living alone

Figure A.2 reproduces Figure 5.3 from the main text using different weights and shows that weight choice makes a difference to the age profile in a few ways. The StatsSA person weight is applied in Panel A and yields the smoothest trends. The StatsSA household weight is applied in Panel B; and in Panel C we apply the cross-entropy weight we constructed in Chapter 3, and which we referred to as CEW1 in Chapter 4. Some discontinuities are observable in these two latter panels which appear related to how the age constraints for household heads entered the calibration. The StatsSA Headship Model used to create the household weight for the GHS calculated headship rates using four age categories: 0-34 years; 35-49 years; 50-64 years; and, 65 years and older. This model (and these age categories) also enter in the calibration of CEW1 and updates to this weight, including the Best CEW. Discontinuities in Panels B and C correspond to these changes in age category, indicated by the red vertical lines. Note the discontinuity in Panel B arises only in the GHS (and notably more so in the early GHS), but for the CEW1-weighted panel, they also occur in the OHS. This is because the StatsSA household weight in the OHS is not calibrated on household totals in the way the household weight is in the GHS. Similarly, discontinuities crop up in the OHS in Panel C because the CEW1 imposes household constraints on the

OHS surveys. This exercise shows how influential weighting can be.

Differences also appear in the levels of the profiles. CEW1 is producing a mix of Panels A and B which appears to result in shifting the female age-profile up slightly, especially at younger ages and in the OHS; but shifting the male headship profile down slightly in middle and older age categories. Interestingly though, both the StatsSA person and household weight detect more male headship in the 50-64 age category, for example, than the CEW1. These discontinuities bleed through into other results. We also find that they impact the cohort results and the results for single-person households in Figures 5.4 and 5.7 in the main text. Ultimately and in spite of these discontinuities, Panel C is the least biased of the three panels overall because it was calibrated on more constraints than either the StatsSA person or household weight and uses both person and household information in the process. Finding these discontinuities simply means that work on the reweighting is not finished. Thankfully, the weighting and how age enters the constraints is something we can change, meaning this is something we can finesse, probably with finer age categories, for example.

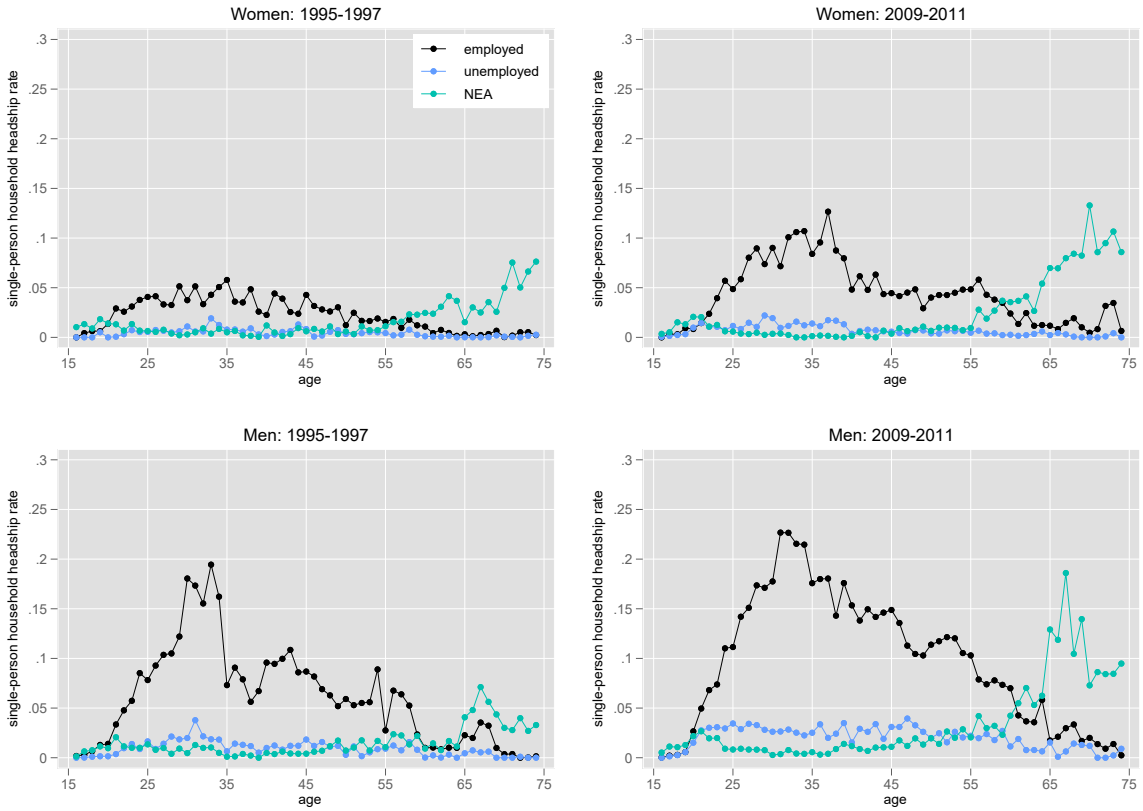
Figure A.2: Age-specific headship rates by age and gender in the OHS-GHS series, differently weighted



Notes: own calculations using the OHS-GHS series. Vertical lines correspond to age categories for StatsSA Headship Model which are 0-34 years; 35-49 years; 50-64 years; and 65 years and older.

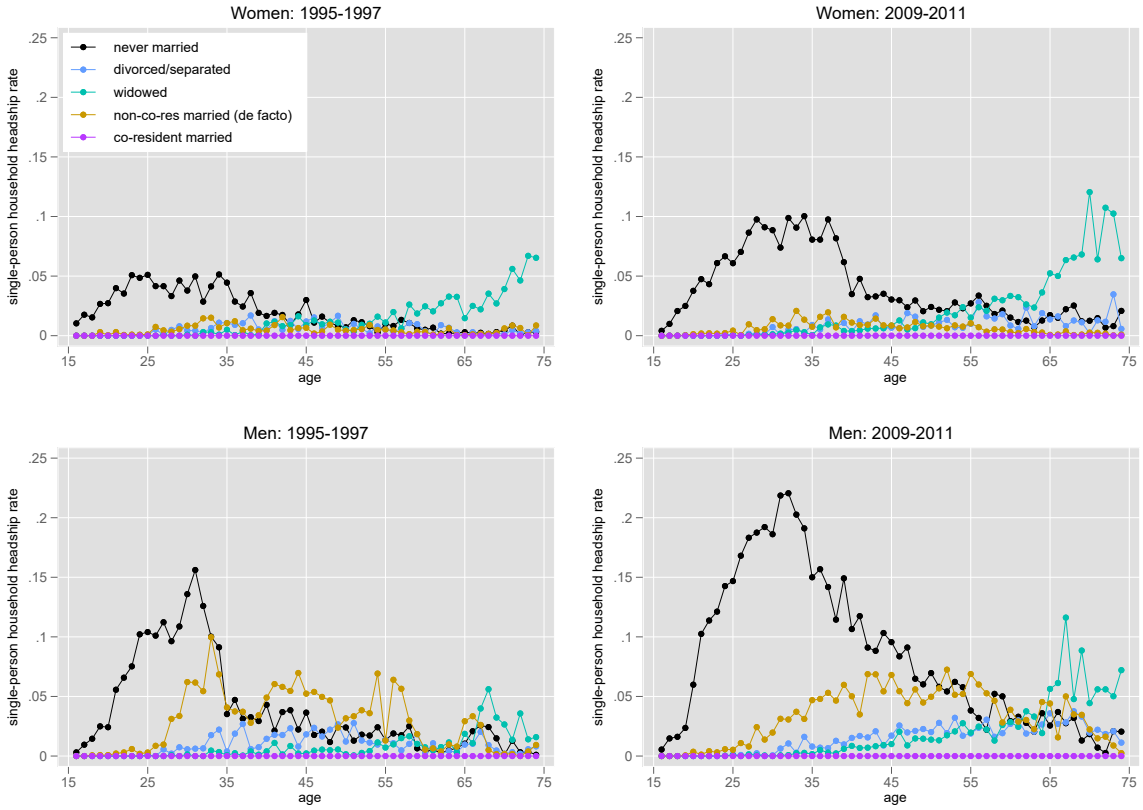
A.3 Labour market, marital status, and living alone in life-cycle perspective

Figure A.3: Living alone and labour market status over the life-cycle by gender, 1995-7 and 2009-11



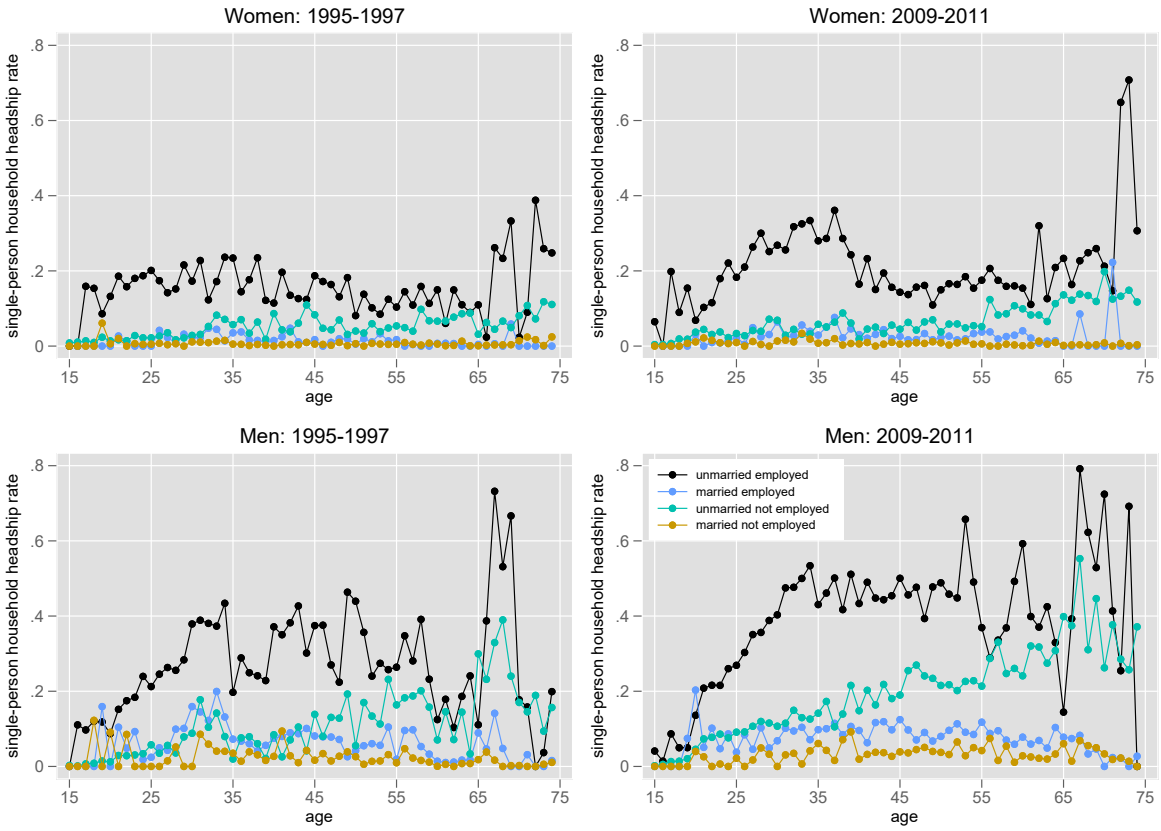
Notes: own calculations using the OHS-GHS series weighted using Best CEW. NEA = not economically active. Unemployed = unemployed by the broad definition. Single-person household headship rate = share of age group both living alone and being of a given labour market status.

Figure A.4: Living alone and marital status over the life-cycle by gender, 1995-7 and 2009-11



Notes: own calculations using the OHS-GHS series weighted using Best CEW. non-co-res married (de facto) = household head is married but not co-residing with a spouse. co-resident married = household head is married and spouse is co-residing. Single-person household headship rate = share of age group being both heads and of a given marital status.

Figure A.5: Single-person household headship rates conditional on labour market and marital status across age by gender in 1995-7 and 2009-11



Notes: own calculations using the OHS-GHS series and weighted using the Best CEW. Not Employed = unemployed or NEA. Unmarried = widowed, divorced/separated, or never married. Married = married or cohabiting..

Appendix B

Additional material for Chapter 6

B.1 Interacting marital and employment status

As explained in the text, marital status has typically been omitted from econometric models of household formation using South African data. In Chapter 6, we not only include marital status, but also it's interaction with employment status because of the endogeneity between these two variables. In the text, we straightaway present our model with the interaction (Model 6.1) and do not take the time to see how the model changes when variables and interactions are added for reasons of space. We go through that process in this appendix by presenting the results of four regression models. Models B.1 and B.2 include only marital status and employment status, respectively. Model B.2 then would be most comparable to models commonly found in the South African literature. Both marital and employment status are included in Model B.3 and both are included plus an interaction in Model B.4.

The notation in the appendix is the same as in Section 6.4.4 in the text; the control variables in X are the same; and the models are run separately by sex. Model B.4 is thus the same as our main Model 6.1 in the text, except that in the case of the former the sample is pooled across time for brevity of presentation. We also only present results for the outcome of headship (H) for the same reason. The following specifications are run for each individual i in a province p:

$$H_{ip} = \beta_0 + \beta_1 \sum_{i=1}^{n} MS_{ip} + \beta_4 X_{ip} + \beta_5 P_p + \epsilon_{ip} \tag{B.1}$$

$$H_{ip} = \beta_0 + \beta_2 E_{ip} + \beta_4 X_{ip} + \beta_5 P_p + \epsilon_{ip} \tag{B.2}$$

$$H_{ip} = \beta_0 + \beta_1 \sum_{i=1}^{n} MS_{ip} + \beta_2 E_{ip} + \beta_4 X_{ip} + \beta_5 P_p + \epsilon_{ip} \tag{B.3}$$

$$H_{ip} = \beta_0 + \beta_1 \sum_{i=1}^{n} MS_{ip} + \beta_2 E_{ip} + \beta_3 \sum_{i=1}^{n} MS_{ip} * E_{ip} + \beta_4 X_{ip} + \beta_5 P_p + \epsilon_{ip} \tag{B.4}$$

The results for these regressions are presented below in Table B.1. The results for Models B.1 and B.2 show that marital status and employment status each appear important when the other is not included. Little changes for women in the size and significance of effects when we move to Model B.3; but effects of both variables weaken for men. Note also that the base group is now married people who are not employed in Models B.3 and B.4. Including an interaction of these variables in Model B.4 is the more notable update for women. Employment makes little difference to the chance of headship for a married woman; but enormously increases the likelihood of headship for a never-married woman. The key insight here is that the effect of employment on household headship varies substantially across marital status; and that this variation is masked in Model B.3 and invisible in Model B.2. These results provide strong empirical motivation for the inclusion of an interaction we previously motivated theoretically in Section 6.2.

Table B.1: The effect of including an interaction in the headship model

YVAR: Headship	**Women**				**Men**			
Model	**B.1**	**B.2**	**B.3**	**B.4**	**B.1**	**B.2**	**B.3**	**B.4**
MS: never married	0.20***		0.20***	0.10***	-0.43***		-0.37***	-0.45***
	(0.00)		(0.00)	(0.00)	(0.01)		(0.01)	(0.01)
MS: divorced/separated	0.52***		0.50***	0.43***	-0.24***		-0.21***	-0.28***
	(0.01)		(0.01)	(0.01)	(0.01)		(0.01)	(0.01)
MS: widowed	0.58***		0.57***	0.54***	-0.11***		-0.08***	-0.12***
	(0.00)		(0.00)	(0.01)	(0.01)		(0.01)	(0.01)
E: employed		0.17***	0.15***	0.02***		0.29***	0.21***	0.13***
		(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.01)
MS*E: never married				0.26***				0.12***
				(0.01)				(0.01)
MS*E: divorced/separated				0.17***				0.12***
				(0.01)				(0.02)
MS*E: widowed				0.10***				0.08***
				(0.01)				(0.02)
Controls + Province FE	Y	Y	Y	Y	Y	Y	Y	Y
_cons	0.23***	0.29***	0.16***	0.21***	0.79***	0.44***	0.64***	0.70***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	320 223	320 223	320 223	320 223	264 826	264 826	264 826	264 826

Notes: own calculations using the OHS-GHS series weighted using the Best CEW on a sample of respondents aged 15 and older. * p<0.05, ** p<0.01, *** p<0.001. Cluster-robust standard errors in parenthesis. Base groups for categorical variables are as follows: MS = married; MS*E = married or not employed. Sample is pooled over time 1995-2011.

B.2 Logit regression model output

Table B.2: Logit regression output for the headship model (marginal effects)

YVAR: Headship	Women			Men		
	OHS 1996-8	Start GHS 2002-4	End GHS 2009-11	OHS 1996-8	Start GHS 2002-4	End GHS 2009-11
MS: never married	-0.0126*	0.0819***	0.146***	-0.306***	-0.275***	-0.269***
	(0.00688)	(0.00650)	(0.00568)	(0.00643)	(0.00668)	(0.00703)
MS: divorced/separated	0.241***	0.269***	0.325***	-0.170***	-0.181***	-0.157***
	(0.00946)	(0.00972)	(0.0125)	(0.0123)	(0.0119)	(0.0158)
MS: widowed	0.301***	0.365***	0.394***	-0.137***	-0.111***	-0.0573***
	(0.00604)	(0.00711)	(0.00706)	(0.0157)	(0.0127)	(0.0173)
E: employed	-0.0202***	0.0427***	0.0289***	0.0917***	0.132***	0.117***
	(0.00611)	(0.00647)	(0.00646)	(0.00637)	(0.00786)	(0.00815)
MS*E: never married	0.225***	0.182***	0.136***	0.0854***	0.0651***	0.0801***
	(0.00911)	(0.00881)	(0.00823)	(0.00903)	(0.0101)	(0.00957)
MS*E: divorced/separated	0.148***	0.136***	0.105***	0.0353*	0.0231	0.0502**
	(0.0156)	(0.0154)	(0.0192)	(0.0205)	(0.0186)	(0.0243)
MS*E: widowed	0.169***	0.106***	0.0895***	0.0537*	0.0512*	0.0764**
	(0.0161)	(0.0163)	(0.0182)	(0.0275)	(0.0275)	(0.0357)
age: 15-24 years	-0.219***	-0.256***	-0.290***	-0.180***	-0.212***	-0.223***
	(0.00778)	(0.00667)	(0.00692)	(0.00630)	(0.00707)	(0.00708)
age: 25-34 years	-0.0706***	-0.0970***	-0.117***	-0.0471***	-0.0607***	-0.0828***
	(0.00479)	(0.00573)	(0.00543)	(0.00578)	(0.00575)	(0.00616)
age: 45-59 years	0.0265***	0.0451***	0.0661***	0.0584***	0.0528***	0.0877***
	(0.00496)	(0.00547)	(0.00543)	(0.00606)	(0.00774)	(0.00687)
age: 60+ years	0.000957	0.0318**	0.0740***	0.0929***	0.119***	0.139***
	(0.00839)	(0.0124)	(0.0123)	(0.0115)	(0.0128)	(0.0171)
pension recipient (aged 60+)	0.0183**	0.0223*	0.0193	0.0601***	0.0284*	0.0776***
	(0.00891)	(0.0122)	(0.0122)	(0.0168)	(0.0158)	(0.0188)
edu: primary school or less	0.00668	0.0190***	0.0229***	0.00308	0.00230	-0.000698
	(0.00431)	(0.00482)	(0.00456)	(0.00456)	(0.00530)	(0.00568)
edu: high school/matric	-0.00342	-0.0113*	-0.0122**	0.0332***	0.00963	0.00472
	(0.00761)	(0.00608)	(0.00572)	(0.00605)	(0.00602)	(0.00537)
edu: post-secondary	0.0248**	0.0435***	0.0618***	0.0803***	0.0786***	0.0952***
	(0.00999)	(0.00899)	(0.00793)	(0.0106)	(0.00980)	(0.0108)
race: coloured	-0.117***	-0.118***	-0.0945***	-0.0310***	-0.0616***	-0.122***
	(0.00933)	(0.0108)	(0.00908)	(0.00984)	(0.00988)	(0.00974)
race: asian/indian	-0.170***	-0.164***	-0.158***	-0.00549	-0.0702***	-0.122***
	(0.0142)	(0.0170)	(0.0163)	(0.0113)	(0.0120)	(0.0137)
race: white	-0.105***	-0.123***	-0.112***	-0.000756	-0.0880***	-0.0940***
	(0.0104)	(0.0115)	(0.0110)	(0.00949)	(0.0110)	(0.0115)
informal dwelling	0.0760***	0.0978***	0.133***	0.108***	0.142***	0.189***
	(0.00638)	(0.00628)	(0.00598)	(0.00820)	(0.00769)	(0.00904)
rural	0.0129**	-9.11e-05	0.000603	0.0337***	-0.00547	-0.00613
	(0.00603)	(0.00610)	(0.00582)	(0.00660)	(0.00718)	(0.00722)
Province FE	YES	YES	YES	YES	YES	YES
N	105 296	109 657	105 269	83 997	93 468	87 360

Notes: own calculations using the OHS-GHS series weighted using the Best CEW on a sample of respondents aged 15 and older. coefficients are marginal effects. * p<0.05, ** p<0.01, *** p<0.001. Cluster-robust standard errors in parenthesis. Base groups for categorical variables are as follows: MS = married; MS*E = married or not employed; age = 35-44 years; edu = incomplete secondary schooling; race = Black African; prov = Western Cape.

Table B.3: Logit regression results for the single-person household model (marginal effects)

YVAR: Single-person household headship	Women			Men		
	OHS 1996-8	**Start GHS 2002-4**	**End GHS 2009-11**	**OHS 1996-8**	**Start GHS 2002-4**	**End GHS 2009-11**
MS: never married	0.0720***	0.0963***	0.112***	0.0675***	0.101***	0.170***
	(0.00631)	(0.00850)	(0.00776)	(0.00971)	(0.00801)	(0.0104)
MS: divorced/separated	0.0997***	0.135***	0.163***	0.195***	0.200***	0.283***
	(0.00639)	(0.00957)	(0.0106)	(0.0115)	(0.0111)	(0.0131)
MS: widowed	0.0917***	0.131***	0.151***	0.177***	0.179***	0.239***
	(0.00637)	(0.00853)	(0.00804)	(0.0135)	(0.0103)	(0.0116)
E: employed	0.0268***	0.0671***	0.0604***	0.0853***	0.111***	0.105***
	(0.00661)	(0.00807)	(0.00837)	(0.0101)	(0.00889)	(0.00998)
MS*E: never married	0.0290***	0.0308***	0.0296***	0.0327***	0.0212**	0.0394***
	(0.00666)	(0.00841)	(0.00893)	(0.00978)	(0.00906)	(0.0105)
MS*E: divorced/separated	-0.00442	-0.0241**	-0.0395***	-0.0264*	-0.0334**	-0.0252
	(0.00825)	(0.0105)	(0.0118)	(0.0145)	(0.0150)	(0.0171)
MS*E: widowed	-0.0127	-0.0431***	-0.0371***	-0.0891***	-0.0584***	-0.0334**
	(0.00807)	(0.00971)	(0.00982)	(0.0181)	(0.0152)	(0.0161)
age: 15-24 years	-0.0317***	-0.0260***	-0.0459***	-0.0660***	-0.112***	-0.142***
	(0.00479)	(0.00458)	(0.00504)	(0.00613)	(0.00595)	(0.00669)
age: 25-34 years	-0.00378	0.00995**	-0.00312	0.0144***	-0.00645	-0.0396***
	(0.00353)	(0.00394)	(0.00394)	(0.00506)	(0.00440)	(0.00521)
age: 45-59 years	-0.00658*	0.00431	-0.0117***	-0.000114	-0.00819*	0.0162***
	(0.00356)	(0.00403)	(0.00429)	(0.00505)	(0.00429)	(0.00543)
age: 60+ years	0.0177***	0.0445***	0.0498***	-0.00557	-0.0210**	0.0240**
	(0.00520)	(0.00572)	(0.00683)	(0.0110)	(0.00833)	(0.0103)
pension recipient (aged 60+)	-0.0299***	-0.0511***	-0.0374***	-0.00545	-0.00911	0.0234**
	(0.00439)	(0.00629)	(0.00689)	(0.0129)	(0.0111)	(0.0115)
edu: primary school or less	-0.00118	0.000980	-0.0174***	0.0162***	0.0103***	9.49e-05
	(0.00308)	(0.00307)	(0.00369)	(0.00410)	(0.00392)	(0.00463)
edu: high school/matric	0.00967**	0.00307	0.00302	0.00916	-0.00864*	-0.00835*
	(0.00447)	(0.00359)	(0.00410)	(0.00567)	(0.00476)	(0.00500)
edu: post-secondary	0.0149***	0.0131**	0.0374***	0.0289***	0.0220***	0.0489***
	(0.00486)	(0.00528)	(0.00453)	(0.00788)	(0.00737)	(0.00818)
race: coloured	-0.0264***	-0.0295***	-0.0492***	-0.0549***	-0.0983***	-0.117***
	(0.00671)	(0.00679)	(0.00678)	(0.0134)	(0.0110)	(0.0108)
race: asian/indian	-0.0247**	-0.0142	-0.00362	-0.0521**	-0.122***	-0.129***
	(0.0124)	(0.0113)	(0.0111)	(0.0225)	(0.0176)	(0.0225)
race: white	0.0368***	0.0550***	0.0468***	-0.0134	-0.0447***	-0.0754***
	(0.00445)	(0.00615)	(0.00559)	(0.00844)	(0.00910)	(0.0114)
informal dwelling	0.0445***	0.0638***	0.0838***	0.108***	0.139***	0.155***
	(0.00435)	(0.00372)	(0.00381)	(0.00730)	(0.00698)	(0.00727)
rural	-0.00952**	-0.0137***	-0.0100**	0.0390***	0.0177**	0.00187
	(0.00459)	(0.00443)	(0.00450)	(0.00713)	(0.00894)	(0.00696)
Province FE	YES	YES	YES	YES	YES	YES
N	**105 296**	**109 657**	**105 269**	**83 997**	**93 468**	**87 360**

Notes: own calculations using the OHS-GHS series weighted using the Best CEW on a sample of respondents aged 15 and older. coefficients are marginal effects. * p<0.05, ** p<0.01, *** p<0.001. Cluster-robust standard errors in parenthesis. Base groups for categorical variables are as follows: MS = married; MS*E = married or not employed; age = 35-44 years; edu = incomplete secondary schooling; race = Black African; prov = Western Cape.

B.3 Detailed Oaxaca-Blinder decomposition output for the linear model

Table B.4: Detailed decomposition output for the headship model

YVAR: HEADSHIP	OHS vs. Start GHS				Start GHS vs. End GHS				OHS vs. End GHS			
OB Effect:	**C**	**W**	**I**	**C+W+I**	**C**	**W**	**I**	**C+W+I**	**C**	**W**	**I**	**C+W+I**
					WOMEN							
MS: never married	4.82	-0.01	0.29	**5.10**	3.71	0.32	0.24	**4.28**	8.32	-0.02	1.08	**9.38**
MS: divsep	0.16	0.14	0.02	**0.32**	0.30	-0.26	-0.05	**-0.01**	0.43	-0.10	-0.03	**0.30**
MS: widowed	0.92	0.84	0.17	**1.92**	0.18	-0.40	-0.01	**-0.23**	1.08	0.50	0.12	**1.69**
E: employed	1.77	-0.06	0.22	**1.93**	-0.53	0.20	-0.08	**-0.41**	1.30	-0.15	0.37	**1.52**
MS*E: never married	0.05	0.53	0.01	**0.59**	-0.73	1.04	-0.24	**0.07**	-0.56	1.55	-0.34	**0.66**
MS*E: divsep	-0.03	0.03	0.00	**0.01**	-0.11	-0.05	0.02	**-0.15**	-0.13	-0.02	0.01	**-0.14**
MS*E: widowed	-0.19	0.11	-0.06	**-0.15**	-0.02	0.00	0.00	**-0.03**	-0.21	0.10	-0.06	**-0.17**
age: < 35 years	-2.57	0.22	0.07	**-2.28**	-2.26	0.18	0.03	**-2.05**	-4.89	0.36	0.20	**-4.33**
age: 45+ years	0.71	0.02	0.02	**0.75**	0.97	0.10	0.05	**1.12**	1.66	0.07	0.15	**1.87**
pension recipient	0.05	0.08	0.02	**0.15**	-0.02	-0.01	0.00	**-0.03**	0.04	0.07	0.01	**0.12**
edu: primary or less	0.53	-0.06	-0.07	**0.40**	0.18	-0.23	-0.05	**-0.10**	0.74	-0.17	-0.27	**0.30**
edu: matric or more	-0.05	0.02	-0.01	**-0.04**	0.15	0.01	0.04	**0.20**	0.07	0.05	0.03	**0.16**
race	-0.22	0.15	0.02	**-0.05**	0.36	0.12	-0.01	**0.47**	0.16	0.25	0.01	**0.42**
informal dwelling	0.41	-0.04	-0.01	**0.36**	0.63	-0.24	-0.09	**0.30**	1.06	-0.22	-0.19	**0.66**
rural	-0.70	0.00	0.00	**-0.70**	0.13	0.03	-0.02	**0.14**	-0.57	-0.06	0.07	**-0.56**
provinces	0.67	0.00	-0.01	**0.66**	0.12	-0.19	0.01	**-0.06**	0.82	-0.23	0.01	**0.60**
_cons	-4.82			**-4.82**	-2.16			**-2.16**	-6.98			**-6.98**
Total	**1.54**	**1.96**	**0.65**	**4.15**	**0.88**	**0.60**	**-0.15**	**1.34**	**2.33**	**1.99**	**1.16**	**5.49**
					MEN							
MS: never married	3.08	-1.28	0.14	**1.95**	3.15	-1.07	0.13	**2.21**	6.10	-2.48	0.55	**4.16**
MS: divsep	-0.05	0.01	0.00	**-0.04**	0.24	0.09	-0.03	**0.29**	0.19	0.09	-0.02	**0.25**
MS: widowed	0.07	-0.05	0.01	**0.03**	0.22	-0.06	0.04	**0.20**	0.27	-0.12	0.09	**0.23**
E: employed	1.61	-0.14	-0.04	**1.43**	-0.96	0.84	-0.13	**-0.25**	0.62	0.50	0.06	**1.18**
MS*E: never married	0.07	0.07	0.00	**0.15**	0.22	0.67	0.09	**0.97**	0.28	0.72	0.13	**1.12**
MS*E: divsep	0.00	-0.01	0.00	**0.00**	-0.03	0.00	0.00	**-0.03**	-0.03	0.00	0.00	**-0.03**
MS*E: widowed	0.00	0.01	0.00	**0.01**	-0.03	0.02	-0.01	**-0.02**	-0.02	0.03	-0.02	**-0.01**
age: < 35 years	-1.76	-0.01	0.00	**-1.77**	-1.75	0.38	0.01	**-1.35**	-3.50	0.31	0.07	**-3.13**
age: 45+ years	0.04	-0.01	-0.01	**0.02**	0.84	0.07	0.03	**0.93**	0.87	0.06	0.03	**0.95**
pension recipient	-0.08	0.08	-0.02	**-0.03**	0.16	0.02	0.01	**0.20**	0.04	0.11	0.02	**0.17**
edu: primary or less	-0.01	0.00	0.00	**-0.01**	-0.12	0.01	0.03	**-0.08**	-0.14	0.01	0.05	**-0.08**
edu: matric or more	-0.43	0.09	-0.08	**-0.42**	0.04	0.21	0.02	**0.27**	-0.37	0.39	-0.16	**-0.14**
race	-1.66	0.01	0.14	**-1.51**	-0.71	0.12	0.00	**-0.59**	-2.38	0.02	0.27	**-2.10**
informal dwelling	0.67	0.05	0.02	**0.73**	1.14	-0.29	-0.12	**0.74**	1.78	-0.17	-0.15	**1.47**
rural	-1.61	0.05	-0.06	**-1.62**	-0.11	0.04	0.01	**-0.05**	-1.71	-0.14	0.18	**-1.67**
provinces	2.62	-0.01	0.02	**2.63**	-0.99	-0.05	-0.04	**-1.09**	1.65	0.03	-0.13	**1.54**
_cons	-2.01			**-2.01**	0.08			**0.08**	-1.93			**-1.93**
Total	**0.55**	**-1.13**	**0.10**	**-0.47**	**1.40**	**0.99**	**0.06**	**2.45**	**1.69**	**-0.67**	**0.96**	**1.98**

Notes: own calculations using the OHS-GHS series weighted using Best CEW on a sample of respondents aged 15 and older. C = coefficient effect; W = endowment effect; I = interaction effect. Effects are the change in conditional probabilities.

Table B.5: Detailed decomposition output for the single-person household model

YVAR: SPHH	OHS vs. Start GHS				Start GHS vs. End GHS				OHS vs. End GHS			
OB Effect:	**C**	**W**	**I**	**C+W+I**	**C**	**W**	**I**	**C+W+I**	**C**	**W**	**I**	**C+W+I**
					WOMEN							
MS: never married	0.14	0.12	0.01	**0.26**	0.58	0.15	0.04	**0.76**	0.68	0.26	0.09	**1.03**
MS: divsep	-0.01	0.03	0.00	**0.02**	0.14	-0.06	-0.02	**0.06**	0.11	-0.02	-0.01	**0.08**
MS: widowed	0.12	0.13	0.02	**0.26**	0.25	-0.06	-0.02	**0.18**	0.33	0.07	0.04	**0.44**
E: employed	0.35	0.01	0.04	**0.40**	-0.31	0.07	-0.04	**-0.28**	0.08	0.02	0.02	**0.12**
MS*E: never married	0.59	0.23	0.12	**0.94**	-0.33	0.68	-0.11	**0.24**	0.32	0.66	0.20	**1.18**
MS*E: divsep	0.02	0.02	0.00	**0.03**	-0.08	-0.03	0.01	**-0.10**	-0.05	-0.01	0.00	**-0.06**
MS*E: widowed	0.01	0.02	0.00	**0.03**	-0.01	0.00	0.00	**-0.01**	0.01	0.02	0.00	**0.02**
age: < 35 years	0.58	0.03	-0.02	**0.59**	-1.00	0.03	0.01	**-0.96**	-0.45	0.06	0.02	**-0.37**
age: 45+ years	0.69	0.00	0.02	**0.71**	-0.13	0.02	-0.02	**-0.14**	0.56	-0.01	0.02	**0.57**
pension recipient	-0.28	-0.10	-0.10	**-0.47**	0.04	0.03	0.00	**0.06**	-0.25	-0.08	-0.08	**-0.41**
edu: primary or less	0.09	0.00	-0.01	**0.08**	-0.63	-0.02	0.17	**-0.48**	-0.64	0.01	0.24	**-0.40**
edu: matric or more	-0.13	0.05	-0.03	**-0.11**	0.28	0.04	0.08	**0.40**	0.10	0.14	0.05	**0.29**
race	0.24	-0.05	-0.02	**0.17**	-0.24	-0.08	0.01	**-0.31**	-0.01	-0.10	-0.02	**-0.14**
informal dwelling	0.44	-0.03	-0.01	**0.40**	0.44	-0.19	-0.07	**0.19**	0.90	-0.15	-0.16	**0.59**
rural	-0.30	0.00	0.00	**-0.30**	0.09	0.07	-0.01	**0.14**	-0.21	0.03	0.03	**-0.16**
provinces	1.32	0.00	-0.02	**1.30**	-1.30	-0.09	-0.02	**-1.41**	0.03	-0.02	-0.12	**-0.11**
_cons	-2.50			**-2.50**	2.46			**2.46**	-0.04			**-0.04**
Total	**1.37**	**0.46**	**0.01**	**1.83**	**0.25**	**0.56**	**0.01**	**0.82**	**1.47**	**0.86**	**0.32**	**2.65**
					MEN							
MS: never married	1.66	0.13	0.08	**1.86**	3.52	0.20	0.15	**3.86**	5.02	0.25	0.45	**5.73**
MS: divsep	-0.10	-0.01	0.00	**-0.11**	0.29	-0.05	-0.03	**0.20**	0.19	-0.07	-0.02	**0.09**
MS: widowed	-0.06	0.04	-0.01	**-0.03**	0.24	0.06	0.04	**0.34**	0.15	0.11	0.05	**0.31**
E: employed	0.88	-0.07	-0.02	**0.78**	-0.85	0.44	-0.11	**-0.52**	0.01	0.26	0.00	**0.26**
MS*E: never married	0.11	0.08	0.01	**0.19**	0.53	0.69	0.21	**1.44**	0.61	0.72	0.28	**1.62**
MS*E: divsep	-0.01	0.00	0.00	**-0.01**	-0.02	0.00	0.00	**-0.02**	-0.02	0.00	0.00	**-0.03**
MS*E: widowed	0.06	0.00	0.01	**0.07**	0.00	0.02	0.00	**0.01**	0.05	-0.01	0.03	**0.08**
age: < 35 years	-2.38	0.00	0.00	**-2.39**	-2.55	0.23	0.03	**-2.29**	-4.93	0.15	0.10	**-4.68**
age: 45+ years	-0.36	0.00	0.00	**-0.36**	0.65	-0.01	0.02	**0.66**	0.29	0.00	0.01	**0.30**
pension recipient	0.00	-0.01	0.00	**0.00**	0.10	0.00	0.01	**0.11**	0.08	-0.01	0.03	**0.11**
edu: primary or less	-0.16	-0.05	0.02	**-0.20**	-0.32	-0.06	0.09	**-0.29**	-0.53	-0.16	0.20	**-0.49**
edu: matric or more	-0.37	0.00	-0.06	**-0.42**	0.31	-0.03	0.08	**0.36**	-0.07	0.06	-0.05	**-0.06**
race	-0.64	0.03	0.03	**-0.58**	-0.51	0.05	0.03	**-0.43**	-1.19	0.05	0.13	**-1.02**
informal dwelling	1.11	0.07	0.03	**1.20**	0.01	-0.42	0.00	**-0.41**	1.12	-0.23	-0.09	**0.79**
rural	-0.80	0.06	-0.03	**-0.77**	-0.88	-0.12	0.12	**-0.88**	-1.65	-0.17	0.17	**-1.64**
provinces	0.96	0.00	0.00	**0.96**	-0.11	-0.05	-0.08	**-0.23**	0.84	0.05	-0.16	**0.73**
_cons	2.08			**2.08**	1.23			**1.23**	3.31			**3.31**
Total	**1.97**	**0.27**	**0.05**	**2.28**	**1.64**	**0.93**	**0.56**	**3.13**	**3.29**	**1.00**	**1.13**	**5.42**

Notes: own calculations using the OHS-GHS series weighted using Best CEW on a sample of respondents aged 15 and older. C = coefficient effect; W = endowment effect; I = interaction effect. Effects are the change in conditional probabilities.

B.4 The Yun (2004) decomposition and detailed decomposition output for the logit model

The *oaxaca* command from Jann (2008) applies the Yun (2004) decomposition when specifying the *logit* option. Sources describing the Yun (2004) decomposition only detail the twofold version of the decomposition, however, the *oaxaca* command is able to run the threefold so the details here are our extrapolation of a threefold version of the decomposition based on twofold examples in Powers et al. (2011), Yun (2004) and Jann (2018). To extract detailed decomposition output, the Yun (2004) decomposition uses a two-step process. The first step is to approximate the gap by evaluating the non-linear functions at the means of their covariates. Let $E(X) = \overline{X_t}$ and $h()$ represent the non-linear link function. Equation 6.5 in the main text is then approximated thus:

$$D \approx [h(\overline{X_1}\beta_0) - h(\overline{X_0}\beta_0)] + [h(\overline{X_0}\beta_1) - h(\overline{X_0}\beta_0)] + [h(\overline{X_1}\beta_1) - h(\overline{X_1}\beta_0) - h(\overline{X_0}\beta_1) + h(\overline{X_0}\beta_0)] \quad \text{(B.5)}$$

The second step is then to linearise this approximation using a first-order Taylor expansion around $\overline{X_0}\beta_0$:

$$D \approx (\overline{X_1} - \overline{X_0})'(\beta_0).f(\overline{X_0}\beta_0) + E(X_0)'(\beta_1 - \beta_0).f(\overline{X_0}\beta_0) + R \quad \text{(B.6)}$$

$$\approx W.f(\overline{X_0}\beta_0) + C.f(\overline{X_0}\beta_0) + R \quad \text{(B.7)}$$

where $f(\overline{X_0}\beta_0) = \frac{dh(\overline{X_0}\beta_0)}{d(\overline{X_0}\beta_0)}$ and W and C are the endowment and coefficient effects, respectively. R corresponds to the interaction (I) term, but residuals from the approximation and linearisation make their way into this term. If the R term is evaluated at a very similar point to the first two terms, point $c \approx \overline{X_0}\beta_0$, then it will approximately be:

$$(\overline{X_1} - \overline{X_0})'(\beta_1 - \beta_0)f(c) \quad \text{(B.8)}$$

which corresponds with the interaction term. This term is better thought of as a residual term to get equation B.7 to balance.

Detailed decomposition output for the coefficient and endowment effects are then arrived at by using the weights from the expansion. Weights (L) for the contribution of covariate k to aggregate effect $s = \{C, W\}$ are then computed as the contribution of that covariate to the the total s effect. Thus the contribution of covariate k to the effect s is calculated as follows:

$$L_s^k = \frac{s^k.f(\overline{X_t}\beta_t)}{s.f(\overline{X_t}\beta_t)} \quad \text{(B.9)}$$

$$= \frac{s^k}{s} \quad \text{(B.10)}$$

These relative contributions are then multiplied by the total s effect to ensure that the detailed contributions add to up to the overall gap. Interpreting detailed decomposition output for the R term is complicated by residuals from the approximation and linearisation. Detailed decomposition can certainly be computed by the *oaxaca* command for the R term, and we report this detailed output, but caution against interpreting these detailed effects too strongly.

Table B.6: Detailed decomposition output for the household headship logit model

YVAR: HEADSHIP	OHS vs. Start GHS				Start GHS vs. End GHS				OHS vs. End GHS			
OB Effect:	**C**	**W**	**I**	**C+W+I**	**C**	**W**	**I**	**C+W+I**	**C**	**W**	**I**	**C+W+I**
					WOMEN							
MS: never married	11.41	-0.05	0.31	**11.68**	4.32	0.36	0.15	**4.83**	16.25	-0.10	1.18	**17.33**
MS: divsep	0.42	0.12	0.02	**0.56**	0.32	-0.23	-0.03	**0.06**	0.83	-0.08	-0.03	**0.72**
MS: widowed	2.23	0.74	0.18	**3.15**	0.47	-0.36	-0.02	**0.10**	2.53	0.43	0.15	**3.10**
E: employed	4.37	-0.09	0.25	**4.52**	-0.57	0.25	-0.04	**-0.36**	2.85	-0.20	0.45	**3.11**
MS*E: never married	-0.74	0.60	-0.07	**-0.20**	-0.71	0.94	-0.13	**0.10**	-1.58	1.70	-0.54	**-0.41**
MS*E: divsep	-0.02	0.04	0.00	**0.02**	-0.09	-0.06	0.01	**-0.14**	-0.15	-0.02	0.01	**-0.17**
MS*E: widowed	-0.26	0.15	-0.04	**-0.16**	-0.06	-0.01	0.00	**-0.07**	-0.29	0.13	-0.05	**-0.21**
age: ¡ 35 years	-6.12	0.31	0.08	**-5.73**	-2.07	0.26	0.02	**-1.79**	-8.61	0.52	0.20	**-7.89**
age: 45+ years	1.88	0.02	0.03	**1.93**	1.21	0.12	0.03	**1.37**	3.50	0.09	0.17	**3.75**
pension recipient	0.10	0.06	0.02	**0.17**	-0.04	-0.01	0.00	**-0.05**	0.04	0.05	0.01	**0.09**
edu: primary or less	1.36	-0.05	-0.08	**1.22**	0.19	-0.24	-0.03	**-0.08**	1.50	-0.13	-0.31	**1.06**
edu: matric or more	0.00	0.03	0.00	**0.03**	0.17	0.04	0.02	**0.23**	0.23	0.07	0.06	**0.36**
race	-1.13	0.20	0.04	**-0.89**	0.52	0.17	-0.01	**0.68**	-0.06	0.34	0.04	**0.31**
informal dwelling	1.04	-0.05	-0.02	**0.98**	0.70	-0.29	-0.06	**0.36**	2.06	-0.27	-0.21	**1.59**
rural	-1.38	0.00	0.00	**-1.38**	0.04	0.00	0.00	**0.04**	-1.10	-0.09	0.08	**-1.11**
provinces	1.14	-0.01	0.00	**1.13**	0.72	-0.25	0.00	**0.47**	2.20	-0.31	-0.01	**1.88**
_cons	-12.88			**-12.88**	-4.40			**-4.40**	-18.01			**-18.01**
Total	**1.43**	**2.03**	**0.69**	**4.15**	**0.74**	**0.68**	**-0.08**	**1.34**	**2.19**	**2.11**	**1.19**	**5.49**
					MEN							
MS: never married	1.99	-1.43	0.39	**0.96**	1.11	-0.88	-0.07	**0.17**	2.56	110.16	0.73	**113.46**
MS: divsep	-0.02	0.01	0.00	**-0.01**	0.06	0.07	0.01	**0.14**	0.05	-3.75	-0.02	**-3.72**
MS: widowed	0.05	-0.06	0.03	**0.02**	0.09	-0.06	-0.02	**0.01**	0.12	6.44	0.13	**6.69**
E: employed	1.70	-0.21	-0.20	**1.29**	-0.76	1.00	0.14	**0.38**	0.47	-29.59	0.15	**-28.97**
MS*E: never married	-0.31	0.11	-0.06	**-0.27**	0.10	0.52	-0.06	**0.56**	-0.14	-42.01	-0.20	**-42.35**
MS*E: divsep	-0.01	0.00	0.00	**-0.02**	0.02	0.00	0.00	**0.02**	0.01	0.10	0.00	**0.11**
MS*E: widowed	0.00	0.01	0.00	**0.01**	0.01	0.01	0.00	**0.02**	0.01	-1.22	0.01	**-1.21**
age: ¡ 35 years	-1.23	-0.02	-0.01	**-1.26**	-0.15	0.50	0.04	**0.39**	-1.07	-23.07	0.02	**-24.12**
age: 45+ years	0.11	-0.01	-0.06	**0.05**	0.40	0.08	-0.02	**0.46**	0.48	-3.64	0.04	**-3.12**
pension recipient	-0.12	0.11	-0.14	**-0.15**	0.15	0.02	-0.02	**0.14**	0.03	-6.38	0.03	**-6.32**
edu: primary or less	-0.03	-0.03	0.02	**-0.04**	-0.07	-0.03	-0.03	**-0.13**	-0.11	3.22	0.13	**3.24**
edu: matric or more	-0.45	0.20	-0.34	**-0.58**	-0.03	0.28	0.00	**0.25**	-0.35	-30.04	-0.46	**-30.85**
race	-1.67	0.01	0.60	**-1.07**	-0.42	0.15	0.00	**-0.27**	-1.68	-0.57	0.60	**-1.65**
informal dwelling	0.56	0.09	0.06	**0.70**	0.42	-0.36	0.06	**0.12**	0.84	11.59	-0.22	**12.21**
rural	-1.50	0.10	-0.26	**-1.66**	-0.01	0.04	0.00	**0.03**	-1.12	9.96	0.38	**9.21**
provinces	2.55	-0.02	0.07	**2.60**	-0.73	-0.08	0.03	**-0.78**	1.17	-1.69	-0.28	**-0.80**
_cons	-1.05			**-1.05**	0.94			**0.94**	0.17			**0.17**
Total	**0.58**	**-1.16**	**0.10**	**-0.47**	**1.12**	**1.27**	**0.06**	**2.45**	**1.44**	**-0.50**	**1.03**	**1.98**

Notes: own calculations using the OHS-GHS series weighted using Best CEW on a sample of respondents aged 15 and older. C = coefficient effect; W = endowment effect; I = interaction effect. Effects are the change in log odds.

Table B.7: Detailed decomposition output for the logit single-person household model

YVAR: SPHH	OHS vs. Start GHS				Start GHS vs. End GHS				OHS vs. End GHS			
OB Effect:	**C**	**W**	**I**	**C+W+I**	**C**	**W**	**I**	**C+W+I**	**C**	**W**	**I**	**C+W+I**
					WOMEN							
MS: never married	0.23	0.18	-0.06	**0.35**	0.15	0.38	0.00	**0.53**	0.47	0.46	0.08	**1.01**
MS: divsep	0.04	0.03	-0.02	**0.05**	0.03	-0.11	0.00	**-0.08**	0.09	-0.03	-0.01	**0.05**
MS: widowed	0.21	0.15	-0.17	**0.19**	0.03	-0.12	0.00	**-0.09**	0.23	0.11	0.03	**0.37**
E: employed	1.47	0.08	-0.83	**0.72**	-0.30	0.36	-0.01	**0.05**	0.75	0.22	0.28	**1.25**
MS*E: never married	-0.11	0.05	0.10	**0.04**	-0.04	0.15	0.00	**0.10**	-0.15	0.18	-0.12	**-0.09**
MS*E: divsep	-0.06	0.00	0.03	**-0.04**	-0.02	0.01	0.00	**-0.01**	-0.09	0.00	0.01	**-0.08**
MS*E: widowed	-0.09	-0.01	0.13	**0.04**	0.02	0.00	0.00	**0.02**	-0.05	-0.01	-0.02	**-0.07**
age: ¡ 35 years	1.40	0.02	0.18	**1.60**	-0.56	0.04	0.00	**-0.52**	0.08	0.06	0.00	**0.14**
age: 45+ years	0.80	0.00	-0.11	**0.69**	-0.19	0.02	-0.01	**-0.18**	0.33	-0.01	0.01	**0.33**
pension recipient	-0.15	-0.07	0.24	**0.02**	0.13	0.02	0.00	**0.15**	0.06	-0.07	0.02	**0.01**
edu: primary or less	0.17	0.01	0.11	**0.28**	-0.43	-0.01	0.04	**-0.40**	-0.82	0.02	0.40	**-0.41**
edu: matric or more	-0.32	0.04	0.27	**-0.01**	0.11	0.06	0.01	**0.18**	-0.10	0.14	-0.07	**-0.03**
race	0.32	-0.03	0.08	**0.37**	-0.20	-0.09	0.00	**-0.28**	-0.13	-0.09	0.03	**-0.19**
informal dwelling	0.15	-0.02	0.02	**0.16**	0.12	-0.17	-0.01	**-0.06**	0.37	-0.13	-0.09	**0.16**
rural	-0.10	0.00	0.00	**-0.10**	0.15	0.10	-0.01	**0.24**	0.21	0.06	-0.04	**0.23**
provinces	1.61	0.00	0.15	**1.76**	-0.64	-0.09	0.00	**-0.73**	0.16	-0.02	-0.14	**0.00**
_cons	-4.31			**-4.31**	1.89			**1.89**	-0.02			**-0.02**
Total	**1.29**	**0.43**	**0.11**	**1.83**	**0.26**	**0.55**	**0.02**	**0.82**	**1.39**	**0.89**	**0.37**	**2.65**
					MEN							
MS: never married	1.80	0.36	0.46	**2.61**	1.66	0.32	0.22	**2.20**	3.37	0.51	0.62	**4.50**
MS: divsep	-0.07	-0.01	0.01	**-0.08**	0.06	-0.07	-0.02	**-0.03**	0.03	-0.09	-0.01	**-0.07**
MS: widowed	-0.06	0.09	-0.04	**-0.01**	0.03	0.09	0.02	**0.14**	-0.01	0.18	0.00	**0.17**
E: employed	0.87	-0.22	-0.13	**0.51**	-0.74	0.83	-0.31	**-0.23**	-0.41	0.58	-0.08	**0.09**
MS*E: never married	-0.32	0.05	-0.08	**-0.35**	0.13	0.17	0.16	**0.46**	-0.06	0.34	-0.06	**0.22**
MS*E: divsep	-0.01	0.00	0.00	**0.00**	0.01	0.00	0.00	**0.01**	0.01	0.00	0.00	**0.01**
MS*E: widowed	0.03	-0.02	0.03	**0.04**	0.01	-0.02	0.01	**0.01**	0.03	-0.04	0.04	**0.04**
age: ¡ 35 years	-2.47	-0.01	-0.02	**-2.50**	-0.64	0.31	-0.05	**-0.39**	-2.59	0.25	0.05	**-2.30**
age: 45+ years	-0.36	0.00	0.01	**-0.34**	0.52	-0.01	0.05	**0.56**	0.44	0.00	0.02	**0.46**
pension recipient	-0.01	-0.01	-0.02	**-0.05**	0.10	-0.01	0.03	**0.12**	0.09	-0.01	0.07	**0.15**
edu: primary or less	-0.42	-0.16	0.30	**-0.29**	-0.24	-0.13	0.23	**-0.15**	-0.67	-0.36	0.51	**-0.51**
edu: matric or more	-0.57	0.06	-0.48	**-0.99**	0.12	0.02	0.11	**0.25**	-0.24	0.20	-0.21	**-0.25**
race	-1.28	0.06	0.39	**-0.84**	-0.12	0.08	0.06	**0.01**	-1.09	0.07	0.27	**-0.75**
informal dwelling	0.40	0.10	0.05	**0.55**	-0.13	-0.35	0.04	**-0.44**	0.11	-0.24	-0.02	**-0.15**
rural	-1.40	0.12	-0.31	**-1.59**	-0.46	-0.13	0.21	**-0.38**	-1.58	-0.24	0.34	**-1.48**
provinces	2.54	0.01	-0.06	**2.50**	-0.84	-0.08	-0.11	**-1.02**	0.66	0.09	-0.31	**0.43**
_cons	3.10			**3.10**	2.01			**2.01**	4.86			**4.86**
Total	**1.76**	**0.42**	**0.10**	**2.28**	**1.46**	**1.02**	**0.65**	**3.13**	**2.96**	**1.22**	**1.24**	**5.42**

Notes: own calculations using the OHS-GHS series weighted using Best CEW on a sample of respondents aged 15 and older. C = coefficient effect; W = endowment effect; I = interaction effect. Effects are the change in log odds.

Appendix C

Additional material for Chapter 8

C.1 Household member type tables including single-person households

Table C.1: Types of household member by gender of the household head, 1995 and 2011: all households

	Male-headed			Female-Headed		
% of households with at least one ...	**1995**	**2011**	**Diff**	**1995**	**2011**	**Diff**
Male-female axis						
Adult of the sex opposite to the head	78.0	60.6	**-17.4**	51.3	49.1	**-2.2**
Spouse	73.2	54.3	**-18.9**	9.6	6.7	**-3.0**
Parent-child axis						
Adult child of the head (aged 15+)	38.2	28.6	**-9.7**	48.4	46.0	**-2.4**
Child of the head (aged <15)	46.0	27.2	**-18.8**	42.4	27.4	**-15.0**
Parent of the head	4.6	1.3	**-3.3**	2.9	1.2	**-1.7**
Extended family						
Adult grandchild of the head (aged 15+)	3.2	4.0	**0.8**	10.0	14.1	**4.1**
Grandchild of the head (aged <15)	13.3	12.5	**-0.8**	26.8	31.5	**4.7**
Grandparent of the head	1.0	0.1	**-0.9**	1.1	0.2	**-0.9**
Brother/Sister of the head	6.5	7.3	**0.8**	10.2	10.3	**0.1**
Other relative of the head	8.3	12.6	**4.3**	12.9	19.1	**6.2**
Other household member	3.0	2.2	**-0.8**	2.0	2.6	**0.6**
Total households (millions)	**5.81**	**8.45**	**2.6**	**3.29**	**5.95**	**2.7**

Notes: own calculations using the OHS-GHS series weighted using the Best CEW. Sample consists of all households.

Table C.2: Types of household members living in households headed by never-married men and women, 1995 and 2011: all households

	Never-married male-headed			Never-married female-headed		
% of households with at least one ...	**1995**	**2011**	**Diff**	**1995**	**2011**	**Diff**
Male-female axis						
Adult of the sex opposite to the head	14.28	9.94	**-4.34**	28.35	32.33	**4.0**
Spouse	0.00	0.00	**0.00**	0.00	0.00	**0.0**
Parent-child axis						
Adult child of the head (aged 15+)	2.39	2.49	**0.10**	22.23	29.15	**6.9**
Child of the head (aged <15)	3.24	1.66	**-1.58**	38.32	31.60	**-6.7**
Parent of the head	5.54	0.66	**-4.88**	4.59	1.00	**-3.6**
Extended family						
Adult grandchild of the head (aged 15+)	0.44	0.68	**0.24**	1.97	4.53	**2.6**
Grandchild of the head (aged <15)	0.95	1.02	**0.07**	9.17	16.11	**6.9**
Grandparent of the head	0.50	0.08	**-0.42**	0.96	0.20	**-0.8**
Brother/Sister of the head	17.48	14.93	**-2.55**	18.90	18.36	**-0.5**
Other relative of the head	8.72	10.91	**2.19**	11.43	19.46	**8.0**
Other household member	5.90	3.15	**-2.76**	2.17	2.69	**0.5**
Total households (millions)	**0.82**	**2.59**	**1.77**	**0.93**	**2.56**	**1.63**

Notes: own calculations using the OHS-GHS series weighted using the Best CEW. Sample consists of all households.