POST-SCHOOL EDUCATION IN AN UNEQUAL SOCIETY

By

Emma Whitelaw

Supervised by

Associate Professor Nicola Branson
and Professor Murray Leibbrandt

A dissertation submitted in fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF CAPE TOWN
School of Economics

Declaration

This thesis is my own work, both in concept and execution. Apart from the normal guidance from my supervisors, I have received no additional assistance.

Signed: Signed by candidate

Emma Whitelaw
April 5, 2023

Acknowledgments

Funding for the research in this dissertation is gratefully acknowledged from the National Research Foundation [NRF] Chair in Poverty and Inequality Research, the Kresge Foundation (via the Siyaphambili Post-school Research Group within the Southern African Labour and Development Research Unit [SALDRU]), and the Spencer Foundation.

Huge appreciation goes to my supervisors, Nicola and Murray, for their kind, gentle, and thorough guidance. Their passion for high-quality, socially meaningful work has been – and will continue to be – an inspiration. I express further gratitude to my colleagues in the Siyaphambili Post-school Research Group and the SALDRU community. To those who have provided feedback, insight, and encouragement along the way, I thank you. This work would not have been possible without helpful discussions, advice, and inspiration from those in the broader post-schooling and education communities in South Africa. I am incredibly thankful for their time. Lastly, I am grateful to my family and friends for their love and support throughout this journey.

POST-SCHOOL EDUCATION IN AN UNEQUAL SOCIETY

Emma Whitelaw
April 5, 2023

Abstract

In South Africa, a country afflicted by conditions of poverty, inequality, and unemployment, a post-school education can be key to fostering upward mobility. However, many of the country's socioeconomic inequalities are replicated within the post-school education system itself. This means that, *inter alia*, inequalities in student access and success plague the sector despite strides made by the government to redress the educational and economic discrimination of the apartheid regime. In the substantive chapters of this dissertation, I explore inequalities in access, academic achievement, and graduate realities; considering each as an obstacle to equitable participation and success in post-school education, and thereafter.

The first substantive chapter concerns access for the 'missing middle'; a group who do not qualify for financial aid but for whom university education is unaffordable. I operationalise the concepts of mobility, vulnerability, and economic stability to differentiate the socioeconomic circumstances of households in South Africa, and locate them within the context of the current post-school funding policy. Results contribute information to an important current policy priority; the development of a sustainable, comprehensive, and progressive financial aid scheme. The second substantive contribution concerns achievement, particularly as it relates to changes in university students' academic performance in 2020 and 2021. Achievement gaps between students funded by the National Student Financial Aid Scheme [NSFAS] and those not funded by NSFAS existed before the COVID-19 pandemic, and the extent to which these were exacerbated by institution closures speaks to issues of participation and success. In my third contribution, I analyse the extent to which financially supporting family and extended family networks is associated with the completion of post-school education. If graduates' realities differ once post-schooling is completed, this can hamper the extent to which post-school education can promote individual upward mobility.

A connecting contribution of these chapters is to provide empirical evidence, through rigorous economic analysis, that builds an understanding of inequalities in access, achievement, and graduate realities. This evidence can be inserted into dialogues that shape policies, which ultimately have the potential to disrupt socioeconomic inequalities. Although concerning different stages of post-schooling, all chapters contribute to quantifying features of post-school education that have not previously been explored in-depth.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction: Post-school education in an unequal society

1.1. The South African post-school sector: A high-level overview

As at the end of 2022, South Africa's post-school education and training [PSET][1] system comprises 26 public Higher Education Institutions (universities hereafter), 131 private universities, 50 Technical and Vocational Education and Training [TVET] colleges, 287 private colleges, and nine Community Education and Training [CET] colleges. Institutions are intentionally differentiated based on their offerings in order to provide for a variety of skills needed in the economy (Department of Higher Education and Training, 2013). For example, TVET colleges offer vocational and occupational education provision, whereas university offerings include both teaching and research activities. Even within the university sector, universities are purposefully differentiated: traditional universities are research-intensive, comprehensive universities have mixed research and teaching focuses, and universities of technology are predominantly teaching institutions.

The largest share of PSET enrolments is in the university sector (65.5% in 2020), followed by the TVET sector (22.6%; 452 227 students) and the CET sector (7.1%; 142 538 students). Out of the total enrolment in the university sector (1 283 890 students), public universities account for the bulk of enrolment at 83.3%, with private universities accounting for the remainder (Department of Higher Education and Training, 2022).[2] The bulk of public spending on post-school education goes to public universities (75.8%; R43.1 billion), followed by TVET colleges (20.7%; R11.8 billion) (Department of Higher Education and Training, 2022).

Over three quarters of students enrolled in public universities in 2020 were Black African students (78.8%; 862 313 students), White students accounted for 10.8% (118 505 students), Coloured students 5.7% (61 923 students), and the lowest enrolment was for Indian/Asian students (3.8%; 41 262 students) (Department of Higher Education and Training, 2021). The Gross Enrolment Ratio [GER][3] for Black African students at public universities, however, sits at

[1] Post-school education refers to any education that takes place after compulsory schooling, outside the general education stream (Department of Higher Education and Training, 2022). Compulsory schooling in South Africa occurs from the year in which a child turns seven until Grade 9 or the age of fifteen, whichever occurs first (South African Schools Act, 1996).

[2] While it is expected that universities will likely meet the National Development Plan [NDP] enrolment target of (1.62 million students) by 2030, the same does not apply to TVET colleges. The average annual growth rate of enrolment at TVETs needs to double per annum from 2019 to 2030 if the NDP enrolment target is to be realised (Department of Higher Education and Training, 2021).

[3] The World Bank Metadata Glossary defines GER as the ratio of total enrolment, regardless of age, to the population of the age group that officially corresponds to the level of education of interest.

just 20% while that of White students is at 49% (Department of Higher Education and Training, 2021).

1.2. Factoring in inequalities

Post-schooling is frequently considered to be a key facilitator of upward mobility and a fundamental component in forging a more equitable society.[4] It is thus not surprising that the ten-year review of the NDP reiterates that "promoting education and skills development remain critical as it is regarded as one of the most effective measures to tackle inequality" (National Planning Commission, 2020, p.28). A focus of South African post-school education policy in the post-apartheid era has thus been to expand access to post-schooling (Department of Higher Education and Training, 2013).

However, economic and educational disadvantages persist in South Africa due to pernicious conditions of poverty, inequality, and unemployment. Consequently, part of expanding of access to post-schooling not only requires places to be made available but needs to ensure that post-schooling is affordable for potential students from a range of socioeconomic backgrounds. To this end, the South African government invests significantly in student financial aid, primarily through the National Student Financial Aid Scheme [NSFAS] (Department of Higher Education and Training, 2013). In the 2022 financial year, approximately R47.3 billion was allocated to NSFAS to fund 691 432 students enrolled at public universities and TVET colleges (South African Government News Agency, 2022).[5]

These funds are distributed to academically eligible students whose combined household income is below R350 000 per annum, which is the reality for approximately 90% of households in the country (Garrod & Wildschut, 2021). Yet, post-schooling is argued to remain unaffordable for an economically vulnerable group, termed the 'missing middle'. An important body of literature on poverty and inequality in South Africa speaks to the concept of economic vulnerability, or instability (see e.g. Carter & May, 2001; Finn & Leibbrandt, 2017; Schotte et al., 2018, 2022; Zizzamia et al., 2019; Zizzamia et al., 2016). In the context of poverty dynamics, vulnerability reflects the idea that poverty affects more households than those that are observed to be poor at a given point in time, because of fluctuations in household income over time. This concept is relevant to student financial aid, since even where household income may render a student ineligible for NSFAS, economic vulnerability or instability may constrain access and/or the ability to pay.

In addition, increased attention has been given to the financial sustainability of the current financial aid model, particularly as budgetary pressures continue to increase in the wake of the

[4]Under a social justice perspective, education is seen to promote "key capabilities that individuals, communities and society in general have reason to value" and to advance poverty alleviation, social mobility, and development (Tikly & Barrett, 2011, p.3).

[5]Only students enrolled at public universities and TVET colleges are eligible for funding through NSFAS.

COVID-19 pandemic. A ministerial task team is currently developing a new student funding model to be introduced in 2023, to address issues of financial sustainability and funding the missing middle (Department of Higher Education and Training, 2020; National Treasury, 2022). In the meantime, an additional R32.6 billion has been allocated to NSFAS over the medium term to continue the current scheme (National Treasury, 2022). Although NSFAS has been key in facilitating access to post-schooling among previously under-served groups (NSFAS, 2016), challenges of relatively low graduation, throughput, and student success rates remain – although these have been improving over time (Department of Higher Education and Training, 2020).

Advancing equitable access and success in post-schooling is complicated by the fact that the post-school system itself can entrench and perpetuate inequalities. For example, systemic inequalities both across and within post-school institution (Culligan, 2022; Department of Higher Education and Training, 2020) persist from historical discrimination of institutions. That is, discrimination under apartheid legislation governed the race group that an institution was designated to serve, and allocated differential resources and funding across institutions.

Inequalities that persist from historical discrimination can perpetuate socioeconomic inequalities today. For instance, admission to post-school institutions remains linked to socioeconomic advantage. Wildschut et al. (2020) find that between 2005 and 2015, the majority of NSFAS-funded (low-income) students were enrolled at historically disadvantaged (Black) institutions [HDIs]. The labour market absorption rates associated with these universities were also significantly lower than historically advantaged (White) institutions [HAIs]. This association holds irrespective of race, gender, and field of study (Wildschut et al., 2020).

Similarly, van Broekhuizen (2016) finds that much of the racially-delineated differentials in graduate employment outcomes can likely be attributed to heterogeneity in the historical type of institutions. Although it is plausible that institutional quality is driving this result, as van Broekhuizen (2016) notes, selection into institutions is endogenous, and graduates from HDIs may be fundamentally different from those who graduate from HAIs.

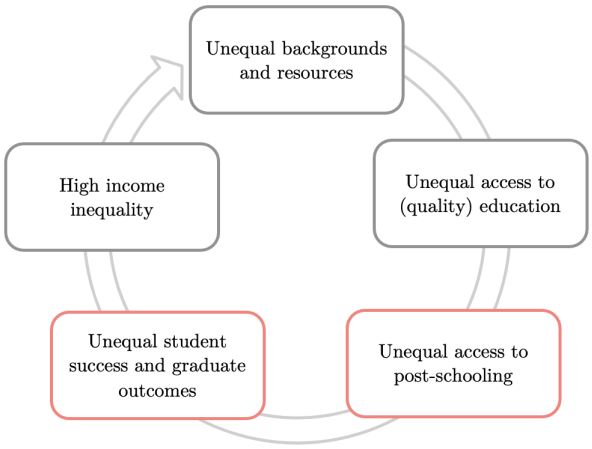
The above research shows that a system that should – and often does – act as a remedy for structural transformation and reform can simultaneously perpetuate cycles of intergenerational social inequality. As Allais et al. (2021) so coherently describe:

> *"Put more bluntly, education is regularly regarded as a panacea for addressing poverty and inequality, despite being bound up in the same reproductive processes caused by poverty and inequality. … Poverty and inequality are the real binding constraints on the ability of the education system to improve and offer greater equality of educational opportunity, and on the ways in which education can reduce income inequality…"*

In this thesis, I aim to contribute to empirically understanding inequalities in post-school education. Across my three substantive chapters, I present economic analyses on topics related

to inequalities in 1) access, 2) achievement, and 3) graduate realities.[6] The role of each of these dimensions in a cycle of inequality (highlighted in orange) can be visualised in Figure 1.1 below.[7] For the sake of brevity, I omit many of the other socioeconomic processes that contribute to this cycle. When inserted into relevant policy dialogues, findings may inform evidence-based decision making that, ultimately, has the potential to disrupt cycles of socioeconomic inequality.

Figure 1.1 The contribution of inequalities in post-schooling to a cycle of inequality



Source: Author's own adaptation from Branson et al. (2020).

Although concerning different stages of the post-school journey, each chapter contributes to quantifying features of post-school education that have not previously been explored in-depth. I acknowledge, however, the growing body of research on inequalities as they relate to post-schooling – both in South Africa and internationally. This research has paved the way for the analyses in this thesis, provided essential context on issues relating to inequalities in post-schooling, and revealed new avenues for relevant and impactful research. This literature will be discussed in detail within each of the following chapters. In the following section, I describe my research agenda, laying out a road map for the thesis, and discussing key motivations for each chapter.

1.3. A post-schooling research agenda

1.3.1. Access

Barriers to entry and participation in the post-school education system remain prevalent in South Africa. While the capacity of the post-school system influences student enrolment (and therefore access), as do inequalities in the quality of South Africa's foundation schooling (Spaull, 2019), evidence suggests that credit constraints are a relevant factor for accessing a post-school

[6] I acknowledge upfront that I do not (and indeed cannot) aim to address all inequalities related to post-schooling in this thesis.

[7] This figure is an adaptation from the Siyaphambili website, also presented in Branson et al. (2020).

education (Branson & Kahn, 2019; Lam et al., 2013; Salisbury, 2016). Despite growth in the scope of NSFAS over recent years, as it currently stands, the financial eligibility threshold on household income has left a group termed the 'missing middle' unable to afford post-schooling and/or accruing large amounts of debt to institutions.[8]

If free education for all students is not viable, a policy that differentiates potential students according to socioeconomic need may be relevant. One possible approach to differentiating socioeconomic need would be to simply consider the household circumstances of those on either side of a household income threshold. However, guided by the poverty dynamics literature, I argue that economic stability – or lack thereof – is a key dimension off which to differentiate households, because it affects the type of decisions that individuals can make. For example, relatively more secure and stable households are typically better able to absorb risk and plan for the future.

Since mobility over time is associated with measurable differences in household characteristics related to economic security and stability, a second approach to differentiating socioeconomic need recognises that mobility patterns, as well as current living standards, are relevant contributions to policy dialogues. In Chapter 2 of this thesis, I thus propose a framework that takes account of mobility around the current NSFAS income eligibility threshold, and not only current income status. This approach recognises that an individual's household income position relative to a threshold can change and is linked to economic stability – or lack thereof. Such an approach allows a profiling of the characteristics of households in a way that could be useful for considering a differentiated approach to funding.

Drawing on data from the National Income Dynamics Study [NIDS] panel, I make two conceptual and methodological contributions through the analysis in this chapter. First, conceptually, the chapter establishes a well-defined framework for the analysis of the missing middle and other eligible groups as they relate to funding thresholds, over time. Second, from a methodological point of view, to the best of my knowledge this is the first study to apply an income dynamics approach to studying funding thresholds, while addressing simultaneously the endogeneity of initial conditions and panel attrition.

1.3.2. Achievement

Even if there is equitable access to post-school education, this does not ensure equitable participation *in* post school education. A stark reminder of this arose during COVID-19-related closures of institutions in 2020. Although by no means the only indicator of a students' post-schooling experience, academic performance is an important measure of student success. Before the pandemic hit, on-campus learning and living typically presented a way – albeit imperfect – of equalising access to resources within an institution for students from differing backgrounds. However, the pandemic suddenly changed the environment (both institutional and

[8]In March 2022, student debt was estimated to stand at R16.5 billion (Gumede, 2022).

economic) on which many students depended to perform academically. That is, campus closures and the move to online learning will have increased the relevance of the home environment to the educational environment.

In particular, key concerns stemming from closures were that many students lacked access to internet, data connectivity and devices, and that glaring structural inequalities shaped the household environment in which many students found themselves, and in which they were expected to learn new academic material. This means that the pandemic may have reinforced – or worsened – pre-existing inequalities in student success. Additionally, changes in academic performance during the pandemic could unintentionally impact academic performance in the long term if students proceed to higher levels without sufficient baseline knowledge. Relatedly, concerns have been raised about the quality of grades and graduation as labour market signals of ability and knowledge acquired. Although inequality in the post-school system may be exacerbated through inequalities in access to and quality of devices, connectivity, and other online infrastructure, on the other hand, the increased use of technology in learning can also offer equalising benefits (e.g. students who are not first-language speakers of English have the opportunity to (re-)listen to lecture recordings at their own pace).

In Chapter 3, I explore changes in university students' academic performance during the pandemic. I use longitudinal administrative data from the University of Cape Town [UCT] to estimate a difference-in-differences model in which the effect of the pandemic on students' performance is allowed to differ by their academic performance in previous years, as well as their socioeconomic status (proxied by NSFAS funding). Thereafter, I further examine academic performance in 2021, considering that changes in performance during the pandemic may have affected performance in following years.[9]

1.3.3. Graduate realities

Currently, the majority of the empirical evidence available on the returns to post-school qualifications relates to labour market outcomes. This presents only a narrow view of how post-school education can benefit both individuals and society. Relevant to positioning the post-school sector as a tool for future mobility and reform is thus broadening and re-framing how the benefits and social externalities of post-school education are conceptualised. Consider, for example, the enrolment of a group of high-income students who come from backgrounds of intergenerational wealth, together with the enrolment of a group of students who will be the first in their families tasked with building generational wealth. Considering only the labour market returns of these graduates will conceal the social benefits of post-schooling, as well as inequalities

[9]The study undertaken in this chapter (Chapter 3) falls under the project entitled *Inequalities in Access and Success in Higher Education*, which forms part of the Siyaphambili Project in the Southern Africa Labour and Development Research Unit. Ethics approval for this project has been granted by the Faculty of Commerce, renewable annually on application. Ethics approval for Chapters 2 and 4 is not required, since these studies utilise publicly available data.

in the responsibilities faced by each group to extend support.

In this vein, I argue that graduate realities are a relevant dimension to be considered when building an understanding of post-schooling in a cycle of inequality. In Chapter 4, I use survey data from the most recent wave of NIDS to explore graduate outcomes that extend beyond labour market returns. To this point, I explore whether graduates of post-secondary education face disproportionate responsibilities to make private transfers (i.e. send remittances). That is, if graduate status induces responsibilities to offer support over and above income or employment status, graduates may experience a particular manifestation of what is colloquially termed 'black tax'.

This study, to the best of my knowledge, is the first of its kind to quantitatively analyse private transfers within the context of black tax and the responsibilities associated with the completion of post-school education in South Africa. This study not only contributes to research on returns to post-school education in a developing country context, considering how returns to post-school education are conceptualised as both a private benefit and social externality (see also e.g. Fongwa, 2019), but also to the growing body of research on support practices and patterns in South and Southern Africa through the lens of black tax (e.g. Mangoma & Wilson-Prangley, 2019; Oppel, 2021).

To close this introductory chapter, I reiterate that I do not aim to address all the complexities of disrupting cycles of inequality, and the role of post-schooling therein. Nevertheless, the research presented in this thesis provides rigorous economic analyses taking stock of three important issues in the post-school sector at this crucial junction in time when addressing issues of equity and sustainability are of heightened priority.

Chapter 2

A tale of two thresholds: Social stratification and post-school funding eligibility

Abstract

Questions related to the sustainability of post-school education funding in South Africa, together with issues of expanding access and affordability have been fervently debated over the last decade. In 2018, government announced that the National Student Financial Aid Scheme would fund all eligible post-school students whose annual household income was R350 000 or less. The Department of Higher Education and Training is now focusing on the 'missing middle' – students who come from households whose income is too high to make the R350 000 funding threshold but too low to afford fees. If free education for all students is not viable, a funding instrument that differentiates students according to socioeconomic need may be relevant.

Guided by the poverty dynamics literature, I argue that a key consideration for understanding socioeconomic need – on both sides of the funding threshold – should reflect the household circumstances that generate economic vulnerability, not only household income at a given point in time. Income mobility in particular is associated with measurable differences in household characteristics that are related to economic vulnerability. In this chapter, I conceptualise a stratification schema around the NSFAS funding threshold that is premised on mobility patterns over time as well as current living standards. Household income mobility is estimated using a multivariate probit model that explicitly accounts for endogeneity of initial conditions, unobserved heterogeneity, and non-random panel attrition. Recognising that the majority of post-school enrolment occurs among youth, I then situate this group within my stratification schema. In doing so, I provide a novel input to current discussions about inequalities in access, and the design of a sustainable, comprehensive, and progressive financial aid scheme.

2.1. Background, objectives, and contributions

The South African government invests substantially in expanding access to post-school education for poor and working-class students. Since 2018, funding has more than doubled from R20 billion to R47 billion in 2022. Following a change in funding policy in 2018, financially-eligible students, whose combined household income falls below R350 000 per annum, are awarded full bursaries through the National Student Financial Aid Scheme [NSFAS].[10]

More recent attention has focused on the lack of affordability of post-school education for the 'missing middle' – a group colloquially defined as those whose household income levels are too low for them to afford fees, but too high to meet the NSFAS income threshold for funding eligibility. Missing middle students are thus currently considered those whose household income is between R350 000 and R600 000 per annum. The Department of Higher Education and Training [DHET] aims to increase the number of NSFAS beneficiaries by 43% by 2024, working towards a financial aid system that is inclusive of the missing middle (Department of Higher Education and Training, 2020).

That being said, the sustainability of the present funding model is currently under scrutiny. The design of a sustainable, comprehensive, and progressive financial aid scheme for the future will therefore require a thorough deliberation of the complexities that affect access to post-school education. If free education for all students is not viable, a policy that differentiates students according to socioeconomic may be relevant. Therefore, recognising that there is a distribution of household circumstances both below the current NSFAS threshold, and within the missing middle, is necessary. Accordingly, before policy solutions are remedied, these circumstances should be interrogated and understood.

One of the possible approaches to differentiating socioeconomic need would be to simply consider the household circumstances of those on either side of a household income threshold. However, guided by the poverty dynamics literature, I argue that economic stability – or lack thereof – is a key dimension off which to differentiate households, because it affects the type of decisions that individuals can make. That is, relatively more secure and stable households are better able to absorb risk and plan for the future.

Mobility over time is associated with measurable differences in household characteristics, that are related to dimensions of economic security and stability. A second approach to differentiating socioeconomic need thus recognises that mobility patterns over time as well as current living standards are important for thinking about policy tools. I draw on the work of Schotte et al. (2018) to propose a framework of social stratification that links the demarcation of strata to an analysis of mobility around the current NSFAS household income eligibility threshold (funding threshold hereafter). This approach recognises that an individual's household income position relative to a threshold can change and is linked to economic stability – or lack thereof. This

[10] Alternatively, the applicant is eligible if they are a social grant recipient.

emphasises that vulnerability and economic insecurity are perhaps more important elements in determining economic welfare compared to income status as observed at a single point in time.

Given that a key feature of this framework is its focus on mobility over time, the empirical application requires detailed panel data. I use data from the National Income Dynamics Study [NIDS], which is South Africa's first nationally representative panel study. There are currently five waves of data available, which span the decade 2008-2017. Using the information in NIDS on an individual's household characteristics, I estimate the probability that their household income falls below or moves above the current funding threshold. To do this, I fit a multivariate probit model, which explicitly accounts for non-random attrition in the panel, endogeneity of initial conditions (household income above or below the funding threshold in the base period), as well as unobserved factors that determine both initial position and household income position in the following period. In theory, this allows me to overcome the biases presented by both initial conditions and selective attrition by modelling them jointly with the probability of household income being below the funding threshold in the next period. Using predicted transition propensities from my model, I then define a stratification schema in relation to the funding threshold. I propose five 'classes' of interest that capture an individual's vulnerability to their household income remaining or falling below the funding threshold.

Hereafter, I locate youth (those aged 15-35) within this framework. South Africa is plagued by unprecedented levels of young people not in employment, education, or training [NEETs] (Schirmer & Nkomana, 2021). However, education and training is a promising way of connecting NEETs to the labour market – youth who are least disconnected are those with an accredited post-matric degree or diploma (Schirmer & Nkomana, 2021). A focus on youth circumstances in relation to student funding is thus particularly relevant in a South African context. Furthermore, by considering all youth, I recognise that there are multiple, potentially non-linear, pathways through which individuals may enter post-schooling. Specifically, I address the following research questions:

1. What are individuals' household income mobility patterns across the NSFAS funding threshold, and what determines the chance that household income falls below or remains below the threshold?

2. What does my classification schema contribute to an understanding of household circumstances – particularly of youth – in relation to the current NSFAS funding threshold?

3. How do these findings contextualise the household circumstances of the missing middle?

The remainder of the chapter proceeds as follows. In the following section (Section 2.2), I foreground relevant details of the South African post-school sector and the funding policy environment. Following this, I provide a high-level overview of the poverty dynamics literature off which my analysis is based. This is followed by a detailing of my conceptual framework and

empirical strategy in Section 2.3. Thereafter, I describe the data (Section 2.4). Results are presented in Sections 2.5-2.7. In Section 2.8, I discuss caveats of the approach and the sensitivity of my results. Finally, in Section 2.9, the chapter concludes.

2.2. Related literature

2.2.1. The post-school system and funding policy environment

This section provides a brief overview of the Post-School Education and Training [PSET] system and the operation of NSFAS therein. I first touch on the relevant details of the NSFAS policy that characterised the student funding landscape leading up to the policy change in 2018, before describing the post-2018 policy environment – current discussions around which I hope to contextualise and inform.

The PSET sector comprises 503 registered institutions, of which 26 are public Higher Education Institutions (universities hereafter), 131 are private universities, 50 are Technical and Vocational Education and Training [TVET] colleges, 287 are private colleges and nine are Community Education and Training colleges. Only students enrolled at public universities and TVET colleges are eligible for funding through NSFAS.[11]

Since 1994, the PSET system has made significant progress in expanding access to post-school opportunities (Department of Higher Education and Training, 2020). Headcount enrolments at universities more than doubled from 495 356 in 1994 to 1 074 912 in 2019, while headcount enrolments in TVET colleges increased by 88.2% from 357 885 in 1999 to 673 490 in 2019. However, the Department of Higher Education and Training (2020) notes that to reach the 2030 National Development Plan [NDP] target of a 27% participation rate in universities, the system still requires a substantial increase in enrolment rates over the coming years.

In achieving this, equity in access to PSET programmes remains key: the ramifications of the inequitable distribution of access to and quality of education across racial groups, gender, and geography during apartheid remain as stark inequalities in access and completion in the PSET system to date. That being said, NSFAS has and continues to play a crucial role transforming the student body and considerably broadening access to universities and colleges (NSFAS, 2016).

NSFAS policy before 2018, and the emergence of the missing middle

Before 2018, NSFAS existed primarily as an income-contingent loan scheme, with partial bursary conversions functioning as incentives for students to pass and graduate. NSFAS funds were allocated to institutions according to the institution's disadvantaged students' cost index. This measure took into account the number of African, Coloured and Indian students enrolled at the institution, as well as the full cost of study [FCS] – that is, tuition fees, accommodation, meals, books, and travel, as provided by the institution. The intention was for institutions to

[11] Figure A.1 in Appendix A shows the distribution of public universities and colleges by province.

rank their students based on socioeconomic need (calculated via a means test[12]) and allocate an amount equal to the FCS less an expected family contribution [EFC] until the full NSFAS institutional allocation was disbursed. However, on average, South African students with a combined household income of approximately R122 000 per annum or less were eligible for funding from NSFAS (Commission of Inquiry into Higher Education and Training, 2017).[13]

Under this policy, NSFAS funding was capped at a maximum level. Although increases in the cap occurred at rates higher than inflation, the maximum often remained below the rising FCS, largely due to escalating fees. The NSFAS cap led to students accumulating what is commonly referred to as 'historic debt' – monies owed to institutions by students who could not cover their EFC, or who were underfunded due to the NSFAS cap.[14]

It is against this backdrop that the student-led Fees Must Fall [FMF] protests began in late 2015, which introduced significant volatility into the student-funding landscape. In addition to the NSFAS shortcomings mentioned above, students further stressed that 1) government funding was failing to match increasing enrolments causing roughly a third of eligible candidates to be denied funding (Universities South Africa [USAf], 2016), and 2) the NSFAS cut-off on household income for eligible students failed to reach the missing middle. It was claimed that the issues above were contributing to academically eligible students dropping out or facing financial exclusion due to inability to pay.[15]

It was also argued that the means test required students to 'perform their poverty'. In this regard, the Commission of Inquiry into Higher Education and Training (2017) recorded that students felt humiliated having to convince a system, which they regarded as operating on a process of mistrust, that they were justified in seeking funding. An overwhelming endeavour of the FMF movement was thus to call for government to provide fee-free higher education for all students. This, however, was deemed financially unviable (Commission of Inquiry into Higher Education and Training, 2017; Jacobs et al., 2019). In the end, protests led to a restructuring of NSFAS funding announced in December 2017. This brought about important changes to the

[12]The means test accounted for household income, the number of dependents in the household, and the family's cost of living. This latter measure was used with academic merit to calculate a score on which to rank students and allocate NSFAS funds (SALDRU, 2017). From 2014, NSFAS began rolling out means test waivers for students who were beneficiaries of social grants. They were automatically given a zero EFC as well.

[13]The Commission noted that the R122 000 threshold was premised on the assumption that it represented the upper limit on the lowest band of the personal income tax table, although, it was not revised for a number of years nor was it an official NSFAS policy guideline.

[14]Furthermore, when institutions engaged in a practice known as 'top-slicing' – a distributive process whereby in a response to receiving insufficient funding from NSFAS, institutions disregarded means test results and shared loan amounts among all students who qualified for NSFAS – students were again left underfunded.

[15]This being particularly because institutions do not allow students who owe money to register or collect graduation certificates.

rules of the scheme, which I now turn to describe.

NSFAS policy after 2018, and the ubiquity of the missing middle

NSFAS eligibility for first-time entering new students from 2018 onward was extended to students with combined household income levels up to R350 000 per annum.[16] NSFAS funding now occurs in the form of a bursary, essentially abolishing the previous loan-bursary combination of the previous scheme. The new scheme also no longer caps funding at universities. The full cost of institution-specific tuition and campus accommodation is covered, with nationally standardised limits on transport, books, food, and off-campus accommodation. Students enrolled in state-funded programs at TVET colleges have always been subsidised by the state at 80% of the total program cost and have always been able to apply to NSFAS for a grant to cover the remaining 20% of college fees, plus an allowance, if they are registered for an approved program. The only change for students enrolling at colleges following the NSFAS policy change in 2018 was the increase in the financial eligibility threshold on household income to R350 000 per annum.

Rather than institutions being allocated funds to disburse, funding applications are now handled centrally by NSFAS and allocated to institutions upon registration of an eligible candidate. Students registered at an institution *prior* to 2018 continue on the capped amount of the previous scheme, but with their full loan amount converted to a bursary. The threshold on household income for students who enrolled prior to 2018 and who apply to NSFAS for the first time is now formally set at R122 000 per annum.

The new R350 000 threshold roughly coincides with the mid-point of the third personal income tax bracket at the time (R296 541 – R410 460).[17] However, no motivation was given for this choice of threshold at the time, except that it was extended to include the missing middle. This meant an additional 24% of South African households were covered by the scheme, bringing the total share of households covered to approximately 90% (Garrod & Wildschut, 2021). Garrod and Wildschut (2021) estimate that the share of households in the missing middle who are not covered by NSFAS has now fallen to 6.1% of the population.

These numbers highlight that funding thresholds inherently partition the household income distribution in South Africa, as shown in Figure 2.1. It is apparent that the majority of the distribution falls below the R350 000 threshold, with a small share remaining in the missing middle (between R350 000 and R600 000).[18] While those in the R350 000-R600 000 income

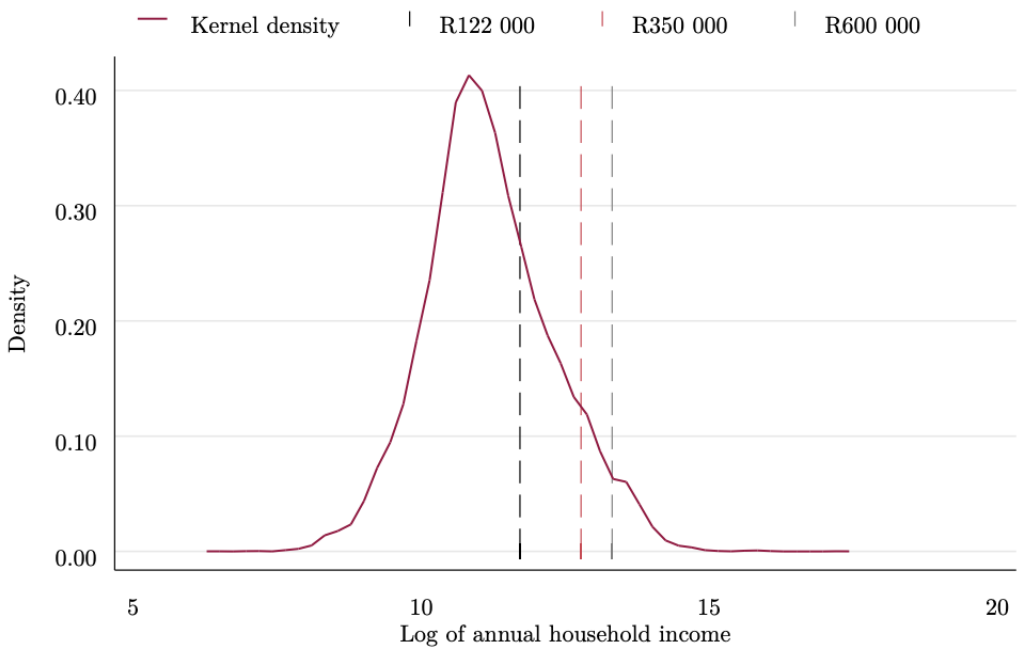
[16]Note that despite using the terminology 'combined household income', since 2018 the formal approach has been to calculate income eligibility only in terms of direct family income (i.e. parents, legal guardians, or spouse). It is not intended to include the income of extended family members if they are not students' legal guardians. Income includes every form of income, both from the formal and informal sector. However, it is not always possible to establish direct family income in the NIDS data. Therefore, total household income is the best measure I have available.

[17]See current and historic income tax rates published by the South African Revenue Service here.

[18]The Commission of Inquiry into Higher Education and Training (2017) notes that the upper limit of

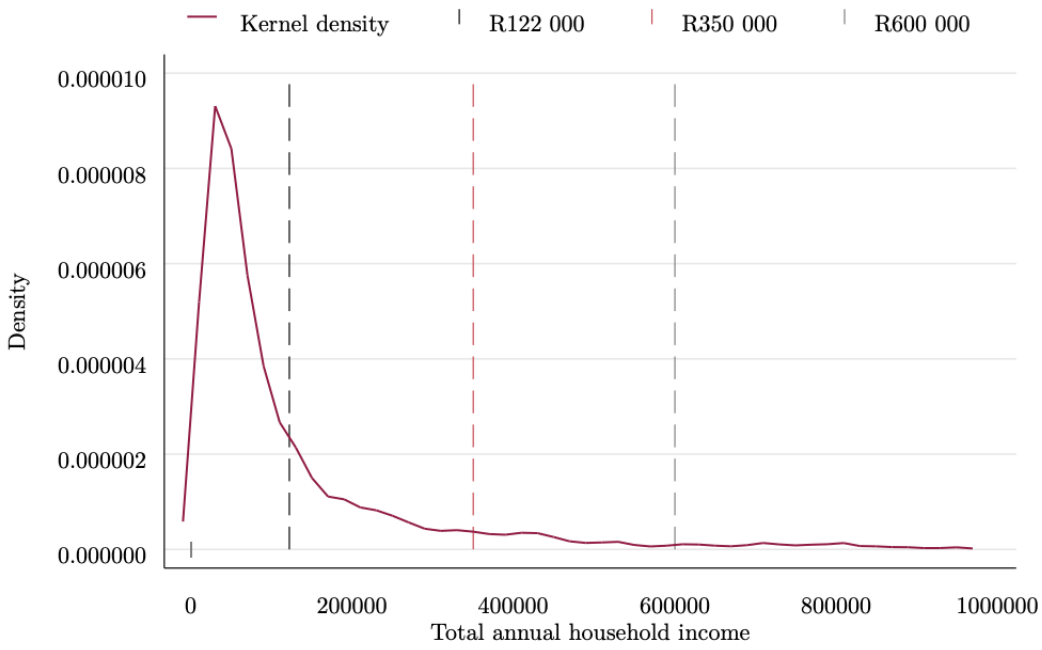
Figure 2.1 Distribution of household income in South Africa, 2017

(a) Annual household income (log) - policy stratification



kernel = epanechnikov, bandwidth = 0.1439

(b) Annual household income (levels)



kernel = epanechnikov, bandwidth = 1.0e+04

Source: Author's own calculations using NIDS Wave 5 (post-stratified weights).
Notes: Distribution in levels is trimmed at the 99th percentile.
One observation per household, income in December 2017 Rands.

bracket may be 'missing' funding, they are in no way in the middle of the household income distribution. This echoes a reflection in Visagie and Posel (2013) that a sizeable share of South Africans in the middle of the income distribution in fact live in poverty.

That being said, this figure does little to situate funding thresholds within conventional notions of the poor, working class ,or middle class – the definitions of which are typically based on per capita expenditure thresholds.[19] For example, the upper bound poverty line, set by Statistics South Africa [Stats SA], was R1 138 monthly expenditure per capita in March 2017 prices (Stats SA, 2019). This does not translate easily to Figure 2.1, which plots annual, household income rather than a monthly, per capita measure. This foregrounds the challenge of situating these funding thresholds in conventional social stratification structures intended to signal socioeconomic disadvantage.

Nevertheless, in the post-school funding discourse, social class definitions based off monetary cut-offs are routinely applied. For example, NSFAS is understood to fund students from *poor and working-class* backgrounds. As evidenced in Garrod and Wildschut (2021), those whose household income is above this threshold are typically considered middle class: "the new scheme does not address the implications of a hard distinction between poor and working-class and *middle class* [emphasis my own]" (p.485). This middle class is often conflated with the missing middle, now considered to be those whose household income is between R350 000 and R600 000 per annum. This perhaps stems from the fact that the concept of the missing middle is not unique to funding policy but stems from economic and political analyses of development (Garrod & Wildschut, 2021). For example, the missing middle has been referenced in the context of South Africa's labour market: "...South Africa has a 'missing middle' of formal sector jobs in the 'middle range' of the formal earnings distribution" (Levy et al., 2014, p.38).[20]

It appears these funding policy thresholds have become conflated with class structures, despite interchangeability of the two being unconvincing. For this reason, although I draw on literature on class dynamics for my analysis, I avoid the terminology of poor, working, and middle class to avoid conflating funding thresholds with poverty and social class thresholds. This is important to keep in mind for the next section, in which I discuss this literature on mobility and social stratification.

2.2.2. Poverty dynamics and class mobility

Literature on the patterns and determinants of poverty in post-apartheid South Africa is well established (see e.g. Carter & May, 2001; Finn & Leibbrandt, 2017; Schotte et al., 2018, 2022;

R600 000 is thought to be premised on an announcement made by the Minister of Higher Education and Training during 2016, in which he pronounced that R600 000 would be used as the upper cut-off for support exempting the poor and missing middle from fee increases the following year.

[19] See Schotte et al. (2018) or Zizzamia et al. (2016) for critiques on the use of absolute monetary thresholds, particularly for defining the middle class.

[20] See also Birdsall (2007) who refers to a 'missing middle' in Africa in the context of development aid.

Zizzamia et al., 2019; Zizzamia et al., 2016). In particular, this literature recognises that households can move in and out of poverty over time. Zizzamia et al. (2019) provide a useful visualisation of this process as a game of snakes and ladders, but with a loaded dice. That is, factors such as family background and geographic location can load the dice in favour of certain individuals. For example, Zizzamia et al. (2019) find that persistent poverty affects primarily African, single-parent, female-headed, and rural households. Higher levels of education of the household head, on the other hand, are strong predictors for lower likelihoods of falling into poverty.

More recently, focus has also fallen on the identification of the middle class. Speaking to social stratification, Schotte et al. (2018, p.89) note that: "In the economic literature, class analyses commonly draw on a monetary indicator to approximate a person's social status. Most commonly used are absolute thresholds that (often arbitrarily) locate the middle class within a particular income or expenditure range, which eases comparisons across countries." However, as the authors point out, the cut-off point separating the poor from the middle class is contested.

For example, Visagie and Posel (2013) define a South African middle class in two ways, and show that the size and composition of the middle class are sensitive to how it is defined – either by the middle share of the national income distribution or an absolute level of affluence – with very little overlap between the two broad definitions. Other authors have suggested that the definition of a middle class should be grounded in economic security. For example, in their study in Chile, Mexico and Peru, López-Calva and Ortiz-Juarez (2014) define the middle class as those who face a maximum risk of falling into poverty of 10%. They state that this is based on empirical evidence[21] that finds that 10% of people in Latin America fell into poverty every year in a 15-year period, using synthetic panels. This points to the existence of an intermediate group of those who can satisfy their basic needs but remain at risk of falling into poverty – this group being distinct from those who can satisfy their basic needs but remain economically stable (Schotte et al., 2018).

This links to the concept of vulnerability, which in this context reflects the idea that poverty affects more households than those that are observed to be poor at a given point in time (Zizzamia et al., 2019). In recognising that households may face non-negligible risks of falling into poverty, Zizzamia et al. (2016) interrogate what this means for the definition of the middle class. They argue that the notion of 'empowerment' is central to both social and political meanings of the middle class. Where a monetary definition of the middle class begins just above a poverty line, the notion of empowerment is at odds with economic instability and vulnerability to poverty. In an attempt to define a middle class that is more appropriately aligned with the idea of empowerment, Zizzamia et al. (2016) present a probability model for social stratification based on the predicted risk of staying in or falling into poverty over a six-year time frame, selecting an

[21] López-Calva and Ortiz-Juarez (2014) cite Cruces, Lanjouw, Lucchetti, Perova, Vakis and Viollaz (2011) here.

expenditure level associated with a maximum risk to poverty of 10% as the lower bound of the middle class (following López-Calva & Ortiz-Juarez, 2014).

Another approach to defining vulnerability follows Cappellari and Jenkins (2004). Cappellari and Jenkins (2004) use panel data from Britain to model poverty transitions and classify households as chronically or transiently poor depending on their chances of exiting poverty or the risk of remaining poor. The authors estimate these transitions using a first-order Markov model, which controls for both initial conditions (those poor in the base year may be a non-random sample) as well as for attrition (panel retention may be non-random). This has been applied in South Africa by Finn and Leibbrandt (2017), who find that ignoring the correlations between the unobservables affecting initial conditions, sample retention, and poverty transitions can lead to substantially biased results.

Schotte et al. (2018) extend Cappellari and Jenkin's (2004) model to above the poverty line, distinguishing a vulnerable group from the stable middle class. Using the model of poverty transitions, the authors predict a person's propensity to remain in or fall into poverty in the near future based on household characteristics and observed poverty status at present. Using conditional propensities of falling into poverty or remaining poor, they then define social classes based on a vulnerability criterion, where vulnerability concerns an above average propensity of falling into poverty.

Schotte et al. (2018) thus propose five social classes – the chronic poor, transient poor, vulnerable, middle class, and elite – where the vulnerable class comprises those who are not poor but find themselves in a position of economic precarity. That is, they have an above average risk of falling into poverty in the near future. Schotte et al. (2018) find that this latter group shares a number of structural commonalities with the transient poor, which would be overlooked in static analyses.

2.3. Analytical and empirical frameworks

2.3.1. Analytical approach

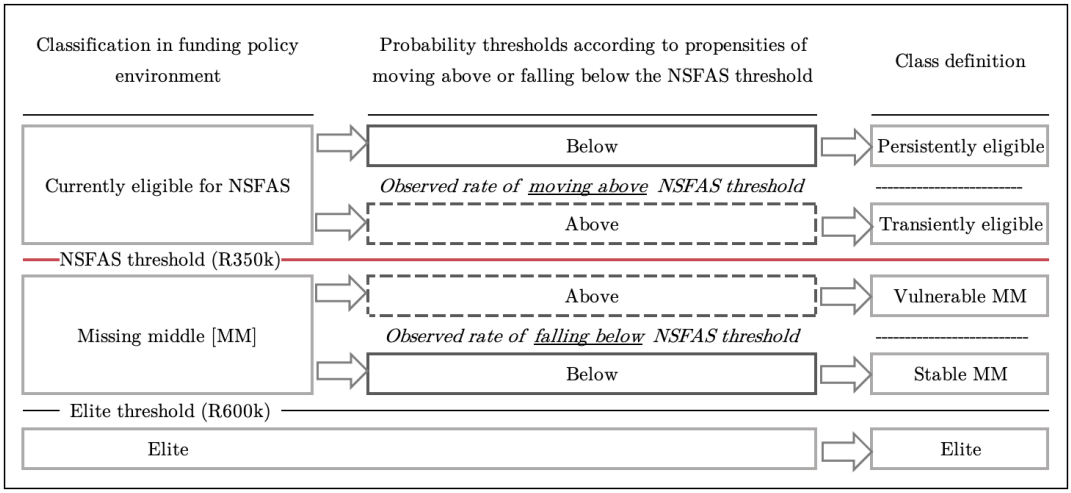
Aiming to enrich knowledge about household circumstances from a perspective that overlooks the dynamics that generate mobility will be fundamentally limiting. In this section, I thus draw on the work of Schotte et al. (2018) to propose a framework for investigating mobility around the current NSFAS funding threshold (i.e. household income moving above or falling below the funding threshold). I propose this will provide a more comprehensive understanding of individuals' and households' medium-term welfare prospects than income alone at a specific point in time.

In Figure 2.2, I illustrate my strategy for classifying individuals based on the probabilities of their household income moving above or falling below the funding threshold. The left column shows conventional class definitions, typically used to reference groups in the post-school funding policy environment. The middle and right columns show how I propose to re-conceptualise

funding classes[22] based on predicted probabilities that an individual's household income moves above or below the threshold in the near future (within two years – the period between survey waves in the data). The empirical strategy for predicting these probabilities is presented in the following section.

I propose five classifications. Those whose household income is below the R350 000 threshold are divided into those who I term 'persistently eligible' and those who I term 'transiently eligible', where the eligibility refers to NSFAS funding eligibility according to household income. Those who are persistently eligible face predicted probabilities below the observed rate of household income moving above R350 000,[23] whereas those who are transiently eligible face predicted probabilities that are above the observed rate of household income exceeding R350 000. The persistently eligible are thus considered to be a relatively more resource-constrained group than those that are transiently eligible. For example, some households may have suffered some negative financial shock that only temporarily pushed their income below the funding threshold.

Figure 2.2 Proposed stratification based on current household circumstances and future mobility



Source: Author's own adaptation from Schotte et al. (2018).
Notes: Observed rate reflects the estimated share of the population to move above or fall below the threshold within two years.

Those in the missing middle are classified as either 'vulnerable missing middle' or 'stable missing middle'. Through using the term 'vulnerable', I hope the capture the idea that households' circumstances can make them vulnerable to falling below a given income level. This is not how economic vulnerability is typically understood, which is related to exposure to exogenous shocks. By using this term, I hope to reinforce the link between the approach

[22] I use the term 'funding classes' to refer to the class structure constructed in my schema, so as not to confuse them with social classes e.g. poor, working class etc.

[23] That is, I estimate the share of the population whose household income was below the threshold in the base period but for whom household income moves above it in the following period, based on observed rates in the data. I elaborate in Section 2.6.

I operationalise here and those which have been used in the poverty dynamics and social stratification literature. I recognise, however, that terms such as 'volatile, 'turbulent', or 'fluctuating' may more appropriately capture the essence of income mobility for this group. I use the term 'stable' to capture the idea of relative non-vulnerability. However, it is important to recognise that while a household may be non-vulnerable, its' circumstances may not ensure it is completely resilient to income falling below the threshold. Those whom I classify as vulnerable have predicted probabilities above the observed rate of household income falling below R350 000. Those whom I term stable have predicted probabilities lower than the observed rate of household income falling below this level in the near future.

It is these observed rates of household income moving above or falling below the funding threshold that I call 'probability thresholds'. This terminology reflects that it is against these rates that I evaluate individuals' predicted probabilities of their household income moving above or falling below the funding threshold. Thus, unlike a threshold against which income is evaluated (an income threshold), my threshold is one against which predicted probabilities are evaluated (a probability threshold). Lastly, the elite are defined as having household income in excess of R600 000 per annum.[24]

A key feature of this framework is its focus on the probabilities associated with household circumstances that generate mobility patterns – vulnerability or stability – over time. Through the implementation of this approach, I hope to communicate a more comprehensive understanding of the socioeconomic circumstances on either side of the current NSFAS threshold in two key ways. First, my approach allows me to compare those currently eligible for funding to those who are not (e.g. how the missing middle compares to those whose household income is below the threshold). This allows me to assess the targeting of the current policy along various dimensions of household well-being. Second, it allows me to provide new context to the missing middle by discerning the extent of heterogeneity in household circumstances between those likely to fall below the NSFAS threshold in the near future, and those who are relatively more secure.

It is important to note that my intention in implementing this approach is not, ultimately, to make a policy recommendation. Rather, I aim to foreground information on the socioeconomic circumstances of South African households in relation to the current threshold, in order to inform and provide context to complex policy considerations going forward.

[24] As a comparison, Schotte et al. (2018) fix their elite threshold (arbitrarily) at two standard deviations above the mean per capita household expenditure. The mean per capita expenditure in their elite group is R19 251 (in January 2015 prices – R22 672 in December 2017 prices), whereas mine is lower at R14 903. Schotte et al.'s (2018) approach contrasts that of Zizzamia et al. (2016), who define the elite cut-off as the average predicted per capita income (or expenditure) of those in the bottom one percentile of the predicted poverty probability distribution. Their elite threshold is set at R10,387 per month (in January 2015 prices – R12 233 in December 2017 prices).

2.3.2. Empirical strategy

Before proceeding, an econometric approach to modelling household income transitions – i.e. the probability that an individual's household income moves above or falls below the NSFAS threshold – is required. My approach follows Cappellari and Jenkins (2004), who introduce the use of a first-order Markov model to examine transitions in and out of poverty, while explicitly controlling for initial conditions effects (those found to be poor in the base year may be a non-random sample) and non-random panel attrition, by jointly estimating current poverty, initial poverty, and sample retention.

Following Cappellari and Jenkins (2004), I implement a first order Markov model to estimate household income transitions around the NSFAS funding threshold by fitting a multivariate probit regression that jointly estimates three equations: 1) the probability of household income being below the funding threshold in time t, 2) the probability of household income being below the funding threshold in time $t-1$, i.e. the initial condition, and 3) the probability of sample retention in time t, where joint estimation allows for correlation between unobservables that influence each of the processes.[25]

The initial conditions equation explicitly accounts for the fact that the point at which I observe an individual in the panel at time $t-1$ does not coincide with that start of the process that generates an outcome in time t (Heckman, 1981). As such, controlling for the observed and unobserved determinants of household income being below the NSFAS threshold time $t-1$ ensures the estimates of transition determinants are less likely to be biased by effects of the covariates on initial conditions. Similarly, non-random panel attrition would lead to a bias in estimates of transition determinants if there is differential response based on transition propensities. The Markovian model is beneficial as it can simultaneously account for initial condition effects, unobserved heterogeneity, and non-random panel attrition in a relatively straightforward way.

While the more traditionally adopted hazard model of income dynamics can also account for unobserved heterogeneity, the ability to account for endogeneity of panel retention and initial conditions is a feature not readily incorporated in the hazard model (Jenkins, 2011). Moreover, Jenkins (2011) notes that in virtually all studies based on hazard regression models, left-censored observations are dropped from the analysis. In other words, observations for respondents whose household income is below the funding threshold in every period would be dropped. This increases the probability that the estimation sample is unrepresentative.

The Markovian model is able to use all available observations for individuals, because it models transitions between one period and the next. In other words, having household income below the funding threshold in this period depends on household income in the previous period, but not further back in time. This contrasts the hazard model, where transitions are duration

[25] Joint estimation is more efficient than separate probits in the case that error terms are correlated (Wooldridge, 2010).

dependent. In this case, transition probabilities would vary with the number of periods someone's household income has been above or below a funding threshold. This means, however, that the dynamics of the Markovian model are not as sophisticated as those inherent to hazard models (Jenkins, 2011). In the Markovian model, the effects of the past arise through 'state dependence', rather than a longer history. Because feedback effects between the previous period and the period before that (e.g. $t-2$) are not examined, the Markovian model is inefficient relative to a model that does exploit these effects. However, the consequences of ignoring the endogeneity of initial conditions and sample retention can substantially bias estimates (Jenkins, 2011).

The Markovian model is specified as follows:

For each individual, $i = 1, ..., N$, define B_{it} to be a binary variable summarising whether or not an individual's annual household income is below the funding threshold in time t. B_{it} is equal to one if an individual's household income in time t is below the threshold, and zero otherwise. Define B_{it-1} to be a binary indicator equal to one if an individual's household income is below the funding threshold in time $t-1$, and zero otherwise. Let R_{it} be a binary variable for panel retention, taking the value one if an individual has household income information in time t and zero if the individual is only observed in time $t-1$. That is, B_{it} is only observed when $R_{it} = 1$. For each time period, individuals can be characterised by the latent propensity of household income falling below the funding threshold, b^*_{it}, latent propensity of initial household income being below the funding threshold, b^*_{it-1}, and latent propensity of sample retention, r^*_{it}, that take the form:

$$b^*_{it} = [(B_{it-1})\gamma_1' + (1 - B_{it-1})\gamma_2']\boldsymbol{x}_{it-1} + u_{it} \tag{2.1}$$

$$b^*_{it-1} = \beta'\boldsymbol{z}_{it-1} + v_{it-1} \tag{2.2}$$

$$r^*_{it} = \psi'\boldsymbol{w}_{it-1} + \varepsilon_{it}, \tag{2.3}$$

where

$$u_{it} = \mu_i + \delta_{it} \sim N(0,1)$$
$$v_{it-1} = o_i + \pi_{it-1} \sim N(0,1)$$
$$\varepsilon_{it} = \eta_i + \xi_{it} \sim N(0,1),$$

and

$$B_{it} = I(b^*_{it} > 0)$$
$$B_{it-1} = I(b^*_{it-1} > 0)$$
$$R_{it} = I(r^*_{it} > 0).$$

The vectors of covariates, $\boldsymbol{x}_{it-1}$, $\boldsymbol{z}_{it-1}$ and $\boldsymbol{w}_{it-1}$, characterise individual i's household in their base year[26] (assumed to be pre-determined); γ, β, and ψ are parameter vectors and u_{it}, v_{it}, and

[26]The use of base-year values better ensures that transitions reflect changes in attributes rather than vice versa.

ε_{it} are the three error terms, assumed to be distributed trivariate standard normal and defined as the sum of an individual specific effect (μ_i, o_i, η_i) and an orthogonal white noise error (δ_{it}, π_{it-1}, ξ_{it}) that follows a normal distribution. The $I(\cdot)$ are binary indicator functions equal to one if the underlying latent propensity exceeds some unobserved value (which can be zero without loss of generality) and equal to zero otherwise.

Specification 2.1 allows not only for base year characteristics to impact income position in the next period, but also for characteristics to have a differing impact by whether the individual's household income fell above or below the funding threshold in the base period – that is, a different impact on the probability of their household income remaining or falling below the funding threshold.[27] Additionally, equation 2.3 can be viewed as a selection equation: it governs whether or not individuals contribute to the estimation of the transition probabilities (γ_1 and γ_2) in equation 2.1.

The log-likelihood contribution of individual i, whose household income is observed in the initial period, is given by the following sample log-likelihood function:

$$\begin{aligned} logL_i =& B_{it-1}R_{it}log[\Phi_3(k_i\gamma_1'\boldsymbol{x}_{it-1}, m_i\psi'\boldsymbol{w}_{it-1}, q_i\beta'\boldsymbol{z}_{it-1}; k_im_i\rho_{31}, k_iq_i\rho_{21}; m_iq_i\rho_{32})] \\ &+ (1-B_{it-1})R_{it}log[\Phi_3(k_i\gamma_2'\boldsymbol{x}_{it-1}, m_i\psi'\boldsymbol{w}_{it-1}, q_i\beta'\boldsymbol{z}_{it-1}; k_im_i\rho_{31}, k_iq_i\rho_{21}; m_iq_i\rho_{32})] \\ &+ (1-R_{it})log[\Phi_2(m_i\psi'\boldsymbol{w}_{it-1}, q_i\beta'\boldsymbol{z}_{it-1}; m_iq_i\rho_{32})], \end{aligned}$$

where $k_i \equiv 2B_{it}-1$, $m_i \equiv 2R_{it-1}-1$, and $q_i \equiv 2B_{it-1}-1$. $\Phi_3(\cdot)$ and $\Phi_2(\cdot)$ are the cumulative density functions of the trivariate and bivariate standard normal distributions respectively. The first term of the log-likelihood function corresponds to the contribution of someone whose household income was below the funding threshold in the base period and who was retained in the sample. The second term is the contribution of an individual whose household income was above the funding threshold in the initial period and who was retained in the sample. The third term is the contribution of an individual whose household income was observed in the initial period, but whose household income was not observed in the following wave (i.e. the individual and/or their household attrited from the sample). However, the use of base year characteristics as covariates means that transition probabilities can be predicted for the attritor subsample (individuals with $R_{it}=0$), using estimates that are robust to non-random attrition.

There are three correlations that are estimated to parameterise the unobserved heterogeneity in the model. These are:

$$\rho_{21} \equiv corr(u_{it}, v_{it-1}) = cov(\mu_i, o_i),$$

which gives the association between unobservable characteristics determining whether an individual's household income is below the funding threshold in time $t-1$ and below the funding threshold in time t,

$$\rho_{31} \equiv corr(u_{it}, \varepsilon_{it}) = cov(\mu_i, \eta_i),$$

[27] Thus while equation 2.1 is an equation for conditional current household income status in relation to the funding threshold, it is convenient to also refer to it as an equation for transitions (Jenkins, 2011).

which gives the relationship between unobservable characteristics determining sample retention and the probability of household income being below the funding threshold in time t, and

$$\rho_{32} \equiv corr(v_{it-1}, \varepsilon_{it}) = cov(o_i, \eta_i),$$

which gives the relationship between unobserved heterogeneity determining sample retention and the probability of an individual's household income being below the funding threshold in time $t-1$.

A test of no correlation between these cross-equation error terms may allow for a simplified model. Other things equal, if $\rho_{21} = \rho_{32} = 0$, there is no initial conditions problem (i.e. initial conditions are exogenous) and having household income below the funding threshold in time t could be simultaneously modelled with sample retention by a bivariate probit regression. Similarly, if $\rho_{31} = \rho_{32} = 0$, then the process governing panel attrition is exogenous. In this case, household income being below the funding threshold in time t could be simultaneously modelled with the initial conditions by a bivariate probit regression. Finally, if $\rho_{21} = \rho_{31} = \rho_{32} = 0$, then both initial conditions and retention are exogenous and household income position in relation to the funding threshold in time t could be estimated by a univariate probit regression.

If $\rho_{21} = \rho_{31} = \rho_{32} = 0$ cannot be rejected, then statistical identification of the model parameters, γ, β, and ψ, requires exclusion restrictions. Specifically, I need to find a set of variables that affect initial state (in time $t-1$) but have no direct effect on household income being above or below the funding threshold in the following period (time t).[28] Similarly, I need to find variables that affect sample retention but have no direct effect on household income being above or below the funding threshold in time t. For the restrictions to be valid, they need to be relevant in the initial conditions and retention equations but excludable from equation 2.1. I test these assumptions in Table 2.3.

Categories of father's occupation (including a category for this information being missing; see Table A.2) satisfy these conditions in initial conditions equation 2.2. I also include parental education and mother's occupation categories in the regression, but these variables do not satisfy excludability from equation 2.1 and therefore are not used as instruments. An indicator variable for whether the respondent was classified by their interviewer as very friendly during the interview satisfies the exclusion restriction requirements for the retention equation 2.3. Here, I assume that friendliness during the interview has no direct effect on household income being above or below the funding threshold once its effect on sample retention has been accounted for. The regression also includes a binary variable for whether or not the respondent was a continuing sample member [CSM] or a temporary sample member [TSM] (as in Cappellari & Jenkins, 2003; Schotte et al., 2018) and a binary variable for attentiveness during the interview, but these variables do not satisfy excludability from equation 2.1 and hence are not used as instruments.[29]

[28] I.e. the variable(s) must be exogenous and excluded from equation 2.1.

[29] It may not be immediately apparent that attentiveness should be correlated with income in time t and

The covariates – which are chosen following previous literature on poverty dynamics and structural well-being (e.g. Cappellari & Jenkins, 2004; Carter & May, 2001; Schotte et al., 2018) – consist of a core set of variables in $\boldsymbol{x}_{it-1}$ expected to be predictors of household income in the South African context (e.g. characteristics of the household head including gender, race, education, age, and employment, together with characteristics of the household including number of residents employed, location of the household, access to certain types of infrastructure, number of durable assets owned, and access to other types of assets, as well as a constant term), with $\boldsymbol{z}_{it-1}$ containing $\boldsymbol{x}_{it-1}$ plus categories of parental education and occupation, and $\boldsymbol{w}_{it-1}$ including $\boldsymbol{x}_{it-1}$ plus indicators for CSM vs. TSM, attentiveness, and friendliness.

Following Cappellari and Jenkins (2004), I estimate the trivariate standard normal distribution function using simulated pseudo maximum likelihood estimation methods based on the Geweke, Hajivassiliou and Keane [GHK] simulator (Gourieroux & Monfort, 1996) with 251 Halton draws.[30] In Section 2.5, I test the exogeneity of the equations 2.2 and 2.3, as well as the exclusion restrictions defined above. I turn now to describe my data.

2.4. Data

I fit the model to data from the National Income Dynamics Study [NIDS] (Southern Africa Labour and Development Research Unit [SALDRU], 2018), the first nationally representative panel study in South Africa. The first wave of data was collected in 2008 on a sample of over 28 000 individuals in about 7 300 households across the country. Individuals from the baseline survey were then recontacted every two years and interviewed along with their current household residents. Along with individually-administered questionnaires, a household questionnaire was completed by the oldest woman in the household and/or another member who was knowledgeable about the household. The NIDS data therefore contain a wealth of information that can be drawn upon to evaluate mobility over time.

Data from pairs of consecutive waves are pooled (i.e. wave 1 to wave 2, wave 2 to wave 3, etc.), and are treated identically in the analysis, controlling for period-specific fixed effects. Data from the first wave (time $t-1$) comprises base year background information, with data from the consecutive wave (time t) indicating whether or not the individual's household income fell above or below the funding threshold. There are currently five waves of data available. My sample thus comprises four pairs of consecutive waves, with an individual contributing up to four pairs.

friendliness not. We know, however, that attrition is strongly correlated with income in our data, so it could be that if high-income individuals do respond in a subsequent wave, they are more likely to be attentive than not. On the other hand individuals could be friendly but still not respond.

[30] This is implemented in STATA 2016 using Roodman's (2011) user-written command `cmp`. Train (2003) notes that Halton draws are more effective than random draws for simulation – that is, a reduction in simulation bias is achieved with fewer Halton draws than random draws. I settle on 251 draws, since the coefficient estimates are stable and practically very similar when I increase or decrease the number of draws to the next prime number.

This results in an estimation sample of 116 462 observations from 46 034 individuals. Note that an individual who does not attrit between waves can contribute the same information in time t in the wave 1 to wave 2 pooling as time $t-1$ information in the wave 2 to wave 3 pooling.

Pooling observations also violates the maximum likelihood estimation assumption of independently and identically distributed errors. In order to correct for this, standard errors are clustered at the household level in the wave in which a respondent first appears in the data (see e.g. Cappellari & Jenkins, 2004; Finn & Leibbrandt, 2017). If an individual responds in all fives waves, their cluster is defined as the household identifier in Wave 1. Those who are added to the sample in subsequent waves are allocated a cluster according to the household in which they are first observed.

It is important to reiterate that NIDS is a panel study that tracks individuals. This means that although funding thresholds are defined at the household level, changes in household income over time will be observed at the level of the individual and will not necessarily be identical across those initially from the same household. Total household monthly income has been multiplied by twelve to get an annual estimate for comparison with the funding thresholds. All monetary measures are deflated to December 2017[31] prices using Stats SA's headline consumer price index (Stats SA, 2017).

Table 2.1 shows that roughly 92% of households receive an annual income less than R350 000, the income eligibility threshold of the present NSFAS policy. Currently, individuals living in households with annual income between R350 000 and R600 000 (the missing middle) are of particular policy relevance. Households in this category comprise 4.43% of households in the sample. This is slightly lower than the share estimated by Garrod and Wildschut (2021), which is 6% of households in the country.

Table 2.1 Sample size by funding policy classifications

NSFAS classifications	Share of households
Household income below R122 000	73.12
Household income between R122 000 and R350 000	18.61
Household income between R350 000 and R600 000	4.43
Household income above R600 000	3.84
Total	100

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratified weights from the base period.

Table 2.2 shows transitions into and out of these categories over time. The figures on the diagonal of Table 2.2 indicate the share of individuals who remained within the same household income category. Those below the diagonal were downwardly mobile, whereas those above the diagonal were upwardly mobile.

Residents of households with annual income between R350 000 and R600 000 show

[31]The month the new policy announcement was made.

particularly interesting mobility patterns. While those in the other categories are most likely to remain where they are, those in the R350 000 and R600 000 category seem as likely, indeed marginally more likely, to fall below the R350 000 threshold in the near future. Panel (b) shows that those with higher household income levels are more likely to attrit over time. This highlights the importance of the empirical analysis correcting for sample selection bias.

Table 2.2 Transitions between NSFAS classifications

Household income (time t-1)	Household income (time t)					
	Below R122 000	R122 000-R350 000	R350 000-R600 000	Above R600 000	Missing	Total
(a) Sample not missing income (time t)						
Below R122 000	86.78	12.19	0.82	0.21	-	100
R122 000 to R350 000	33.36	54.48	9.29	2.88	-	100
R350 000 to R600 000	16.33	31.99	29.91	21.76	-	100
Above R600 000	6.11	23.95	24.82	45.12	-	100
Total	71.4	21.54	4.25	2.81	-	100
(b) All respondents						
Below R122 000	73.14	10.28	0.69	0.18	15.72	100
R122 000 to R350 000	27.27	44.54	7.59	2.35	18.24	100
R350 000 to R600 000	11.37	22.27	20.82	15.15	30.38	100
Above R600 000	4.12	16.15	16.74	30.43	32.57	100
Total	58.93	17.78	3.51	2.33	17.45	100

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratified weights from the base period. Table A.1 in Appendix A applies attrition adjusted weights to the shares in panel (a).

2.5. Empirical application

2.5.1. Model specification tests

I begin by testing the exogeneity of initial conditions equation 2.2 and sample retention equation 2.3 with respect to household income being below the funding threshold in time t. That is, I test the separate and joint significance of the correlation coefficients ρ_{21}, ρ_{31}, and ρ_{32}. I also test the validity of my exclusion restrictions by treating the non-linear functional form of the model as sufficient for identification, and using father's occupation and the indicator variable for friendliness as over-identifying restrictions for equations 2.2 and 2.3 respectively. The results of these tests are presented in Table 2.3.

The evidence in panels (a) and (b) suggests that the estimation strategy for household income transitions around the NSFAS funding threshold is appropriate. From panel (a), I observe significant correlation between unobservables affecting initial position in relation to the funding threshold and falling below the funding threshold in time t (ρ_{21}), as well as initial position (above or below) in time $t-1$ and sample retention (ρ_{31}). Moreover, panel (b) shows exogeneity of initial conditions and sample retention are rejected at the 1% significance level, as is joint exogeneity. Both initial conditions and sample retention are thus be considered endogenous to this model.

Table 2.3 Estimates of model correlations and model test statistics

	Estimate	P-value
(a) Correlation coefficients between unobservables		
Base-year status and conditional current status (ρ_{21})	-0.47	0.00
Retention and conditional current status (ρ_{31})	0.31	0.00
Retention and base-year status (ρ_{32})	0.13	0.00
Null hypotheses	Test statistic	P-value
(b) Wald test for exogeneity of selection equations		
Exogeneity of initial conditions ($\rho_{21} = \rho_{32} = 0$)	21.51	0.00
Exogeneity of sample retention ($\rho_{31} = \rho_{32} = 0$)	21.48	0.00
Joint exogeneity ($\rho_{21} = \rho_{31} = \rho_{32} = 0$)	42.67	0.00
(c) Instrument validity		
Exclusion of father's occupation from transition equation (d.f.=10)	17.66	0.13
Exclusion of friendliness indicator from transition equation (d.f.=2)	0.83	0.66
Exclusion of father's occupation & friendliness indicator from transition equation (d.f.=12)	17.12	0.25
Relevance of father's occupation in initial conditions equation (d.f.=5)	26.21	0.00
Relevance of friendliness indicator in retention equation (d.f.=1)	10.41	0.00

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Simulated pseudo maximum likelihood estimation with 251 random draws. Data are weighted using post-stratification weights from the base period. Standard errors are robust to arbitrary levels of intra-household correlation and the presence of repeated observations on the same individual.

Table 2.3 panel (c) shows that the indicator variables for father's occupation are significantly correlated with household income being below the funding threshold in time $t-1$ but are excludable from the main equation. Similarly, being friendly during the interview affects sample retention, but has no association with household income being below the funding threshold in time t. As such, I am confident that these controls allow for the identification of the system of equations.

2.5.2. Effects of covariates on income transitions

Specification 2.1 allows for differing effects of covariates on transitions depending on whether an individual's household income was above or below the funding threshold in the base period. The model implies the following equations for the conditional probability of household income remaining below the funding threshold (s_{it}) and the conditional probability of household income falling below the funding threshold (e_{it}), respectively:

$$s_{it} \equiv Pr(B_{it} = 1 | B_{it-1} = 1) = \frac{\Phi_2(\gamma_1' \boldsymbol{x}_{it-1}, \beta' \boldsymbol{z}_{it-1}, \rho_{21})}{\Phi(\beta' \boldsymbol{z}_{it-1})}$$

$$e_{it} \equiv Pr(B_{it} = 1 | B_{it-1} = 0) = \frac{\Phi_2(\gamma_2' \boldsymbol{x}_{it-1}, -\beta' \boldsymbol{z}_{it-1}, -\rho_{21})}{\Phi(-\beta' \boldsymbol{z}_{it-1})}$$

Average marginal effects [AMEs] of covariates in the model are calculated by obtaining the

marginal effect [ME] of the variable on the conditional probability of household income remaining below or falling below the funding threshold for each observation, then averaging the ME across observations. The ME for binary variables is calculated as the discrete difference implied by a unit change in the variable of interest (i.e. the difference in predicted conditional probabilities when all individuals are set to one and when all are set to zero, holding all other variables constant). For count variables, MEs are the changes in conditional probabilities when a variable increases by unity, holding all other covariates constant. MEs of continuous variables are calculated by inducing an infinitesimal change in the covariate, with all other covariates held constant.

The AMEs for household income remaining and falling below the funding threshold are shown in Table 2.4. Household income remaining below the funding threshold and household income moving above the funding threshold are mutually exclusive events. That is, the probability of household income moving above the funding threshold is exactly $1 - s_{it}$. This implies that any covariate that is predicted to increase (decrease) the likelihood of household income remaining below, will automatically decrease (increase) the likelihood of household income moving above the funding threshold by the same extent. The same applies to the estimated chances of household income falling below the funding threshold and remaining above the funding threshold.

In general, Table 2.4[32] shows that employment, education, and current living conditions are markers of economic resilience and stability. For example, an individual residing in a household with income below the funding threshold in time $t - 1$, but in which the head has a post-school qualification, is on average 6.5 percentage points less likely to remain below the threshold compared to someone residing in a household below the threshold in which the head has incomplete secondary. For those whose household income is above the funding threshold in time $t-1$, residing with a head who has a post-school qualification reduces the likelihood that the individual's household income will fall below the threshold in the next period by a statistically significant 13.1 percentage points, on average.

Additionally, factors such as having more rooms in one's household, a higher share of durables, and access to a flush toilet are associated with a greater likelihood of household income moving above the funding threshold for those below in the base period. Moreover, residents of households in which at least one resident has a pension asset are significantly less likely to have household income remain below the funding threshold, or fall below, in the next period. Residing with a head who is White is also associated with greater economic stability or upward mobility than a head of any other race group.

[32]The results of the initial conditions and retention equations can be found in Table A.2 in Appendix A, along with the full results on the transition equation.

Table 2.4 Coefficient estimates and AMEs of threshold status in time t, conditional on threshold status in the base period

	Remaining below (s_{it})		Falling below (e_{it})	
Base period covariate	Coefficients	AMEs	Coefficients	AMEs
Characteristics of the household head				
Race (base: White)				
African	0.444**	0.065**	0.702***	0.317***
	(0.18)	(0.03)	(0.21)	(0.05)
Coloured	0.315*	0.032**	0.451**	0.136***
	(0.18)	(0.02)	(0.21)	(0.04)
Indian/Asian	0.548**	0.041***	0.054	0.025
	(0.25)	(0.02)	(0.28)	(0.08)
Age of household head	-0.001	-0.000	0.013**	0.003*
	(0.00)	(0.000)	(0.01)	(0.002)
Household head is female (yes=1)	0.105*	0.013**	0.245*	0.088**
	(0.06)	(0.01)	(0.14)	(0.04)
Highest education (base: Incomplete secondary)				
Missing	-0.537*	-0.065	-2.488***	-0.628***
	(0.29)	(0.05)	(0.49)	(0.08)
No schooling	0.263*	0.028***	-0.938**	-0.197
	(0.14)	(0.01)	(0.46)	(0.14)
Primary	0.052	0.009	-0.177	-0.011
	(0.10)	(0.01)	(0.40)	(0.10)
Matric (Grade 12)	-0.332***	-0.039***	-0.056	-0.034
	(0.10)	(0.01)	(0.21)	(0.06)
Post-school qualification	-0.506***	-0.065***	-0.286	-0.131***
	(0.10)	(0.01)	(0.18)	(0.05)
Household characteristics				
Household employment rate (working age adults)	-0.379***	-0.047***	-0.367	-0.164***
	(0.11)	(0.01)	(0.23)	(0.05)
At least one resident has a pension (yes=1)	-0.381***	-0.051***	-0.060	-0.060*
	(0.10)	(0.01)	(0.14)	(0.03)
Households owns livestock (yes=1)	0.288**	0.025***	1.053***	0.201***
	(0.13)	(0.01)	(0.36)	(0.07)
Share of durables owned by household	-1.162***	-0.149***	1.358**	0.079
	(0.38)	(0.03)	(0.55)	(0.14)
Number of rooms in house	-0.038***	-0.005***	-0.043*	-0.018***
	(0.01)	(0.00)	(0.02)	(0.01)
Household has access to flush toilet (yes=1)	-0.220*	-0.025*	-0.161	-0.077
	(0.13)	(0.01)	(0.24)	(0.06)
Number of household residents	-0.064**	-0.007**	0.024	-0.003
	(0.03)	(0.00)	(0.06)	(0.01)
Number of children under 6	0.095	0.010*	0.230*	0.066**
	(0.06)	(0.01)	(0.13)	(0.03)
Number of children aged 6-18	0.011	0.002	0.011	0.007
	(0.04)	(0.00)	(0.09)	(0.02)
Constant	4.433***		-1.72**	
	(0.89)		(0.77)	
Time fixed effects	Yes			
Province fixed effects	Yes			
Pseudo log-likelihood	-123 000 000			
Model chi-squared (d.f.=178)	6 720.48			
Number of clusters	16 782			
Observations	116 462			

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses and are robust to arbitrary levels of intra-household correlation and the presence of repeated observations on the same individual. Data are weighted using post-stratification weights from the base period. The model also contains controls for the household dependency ratio, whether the household head is employed, whether at least one household resident owns financial assets, if a household member owns the dwelling, if the household has access to piped water, if it has access to electricity, and its location (urban, rural, farm). These are omitted from the table. None of the effects of these covariates are statistically different from zero.

On the other hand, households in more precarious positions economically tend to have limited access to infrastructure and assets, and are home to more young children. Households that own livestock also appear to be more prone to income remaining below the funding threshold, or to falling below the threshold in the near future, with the size of this effect considerably larger for those whose household income is above the threshold in the base period.[33]

Although this framework does not allow me to give a causal interpretation to the impact of household circumstances on mobility, it does permit an understanding of the relative importance of particular household characteristics on vulnerability within the context of post-school funding thresholds, and where youth fall therein.

2.6. Structuring stratification around funding thresholds

In this section, I describe the process for constructing a stratification schema around the R350 000 funding threshold. This allows for further, more nuanced interrogation of household circumstances in relation to the funding threshold. To begin, I calculate the share of individuals in the data for whom I observe household income to fall below or move above the funding threshold. That is, using observations from Waves 1 to 5, I estimate the share with household income below the funding threshold in time $t-1$, but who moved above it in time t.[34] An estimated 3.6% of individuals have their household income move above R350 000 in the next period, when it is below R350 000 in the base period. This is the observed rate of household income moving above the funding threshold.

Similarly, I estimate the share of those whose household income was above the funding threshold in time $t-1$, but whose household income fell below it in time t. This is the average probability of falling below the threshold, estimated based on observed rates in the data. Thirty-eight percent of individuals are estimated to have household income fall below R350 000 in the next period, when it is above R350 000 in the base period. This is the observed rate of household income moving above the funding threshold. Panel (a) of Table 2.5 reflects these observed rates, which I call my 'probability thresholds'. It is against these probability thresholds that I evaluate the predicted conditional probabilities of household income moving above or falling below the funding threshold. These predicted conditional probabilities are those predicted using coefficient estimates from the multivariate probit model.

[33] Although livestock is an asset, 85% of households that own livestock reside in a rural location. The effect may thus be picking up the economic precarity that comes with residing in a rural location.

[34] This follows Schotte et al. (2018) and López-Calva and Ortiz-Juarez (2014) who both set a lower threshold of the middle class to a probability of becoming poor, based on empirical evidence.

Table 2.5 Probability thresholds

	(a) Probability threshold (%)				(b) Associated income threshold			
	Mean	S.E.	95% C.I.		Mean	S.E.	95% C.I.	
Observed rate of **moving above** for those below in the last period	3.60	0.14	3.33	3.88	88 342	1 744	84 922	91 762
Observed rate of **falling below** for those above in the last period	37.79	1.48	34.89	40.69	465 280	6 334	452 803	477 757

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Probabilities are weighted using post-stratification weights from the base period and corrected for panel attrition. S.E. abbreviates standard error and C.I. abbreviates confidence interval. Annual household income is in Rands.

In calculating these observed rates of moving above or falling below the funding threshold, I apply attrition adjusted weights. This is done because not all respondents who are observed in time $t-1$ are re-observed in time t. In order to accurately compute the observed share of those whose household income moved above or fell below the funding threshold in time t, I account for the fact that not all individuals' household income is observed in the next period.[35] Attrition adjusted weights are computed as the product of the post-stratified weight in time $t-1$ and the inverse of the conditional probability of re-interview[36] in time t.

For comparative purposes, I also compute an income value associated with the probability thresholds. These are shown in panel (b) of Table 2.5. The monetary values associated with the probability thresholds are computed as the average annual household income of those whose predicted conditional probability of moving above or falling below the funding threshold is within the 95% confidence interval around the respective probability threshold. On average, those with an annual household income of R88 342 fall into the 95% confidence interval for household income moving above the funding threshold. Those with an annual household income of R465 280 fall into the 95% confidence interval for household income falling below the funding threshold.[37]

I demonstrate the probability thresholds on Figure 2.3, an updated version of Figure 2.2. The observed rate of household income moving above the funding threshold (3.6%) is set as the cut-off point separating those persistently below the funding threshold from the transiently below the funding threshold. Correspondingly, the observed rate of household income falling below the funding threshold for those who were initially above it was 37.79% in the pooled sample. This

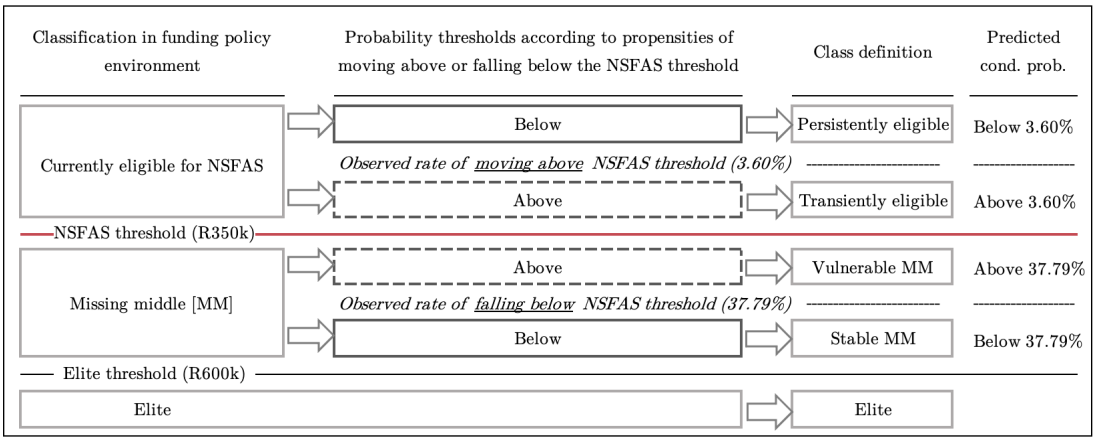
[35] As mentioned, however, the multivariate probit model adjusts directly for attrition. This means I can still predict conditional probabilities of household income moving above or falling below the funding threshold ($1-s_{it}$ and e_{it}, respectively), even for those who attrit. As such, my class structure is defined for all respondents in the sample – even those who attrit in time t. For this reason, attrition adjusted weights are not required throughout the analysis.

[36] The conditional probability of re-interview is estimated using the coefficient estimates from the retention equation 2.3 in the multivariate probit model.

[37] Put differently, R465 280 can be viewed as the value at which individuals are typically able to afford the cost of insuring against their household income falling below R350 000 per annum (Zizzamia et al., 2016).

is set as the cut-off point separating the vulnerable missing middle from the stable.

Figure 2.3 Proposed stratification based on current household circumstances and future mobility (updated with associated income thresholds)



Source: Author's own adaptation from Schotte et al. (2018).
Notes: Cond. prob. abbreviates conditional probability.
Observed rate reflects the estimated share of the population to move above or fall below the threshold within two years.

Finally, I evaluate the performance of my probability threshold in Table 2.6. Table 2.6 identifies the share of those who I identify as vulnerable whose household income has actually fallen below the funding threshold in the next period, and the share of those who I identify as transiently eligible whose household income has moved above the funding threshold in the next period.

It is apparent that the probability threshold performs well throughout, but especially with respect to the vulnerable group where over two thirds of those identified as vulnerable had household income fall below the R350 000 threshold in two years' time. Surprisingly, a non-negligible share of the elite also had household income fall below the R350 000 threshold in the next period. Indeed, those in the upper two classes may still face economic precarity that is not accounted for in my model. To gain a deeper understanding of this stratification schema and its relevance for informing post-school funding policy, I turn now to provide a profiling of household characteristics by funding class.[38] To comprehensively inform policy, it is relevant not only to know who the vulnerable are, but the sources of their vulnerability too.

[38]By construction, characteristics closely resemble the transition determinants reported in Table 2.4.

Table 2.6 Population share and mobility patterns by funding class

	Population	Share that fell below	Share that moved above
Persistently eligible	60.56	-	0.88
Transiently eligible	31.09	-	9.74
Vulnerable missing middle	2.17	67.70	-
Stable missing middle	2.41	29.91	-
Elite	3.76	26.57	-
Total	100		

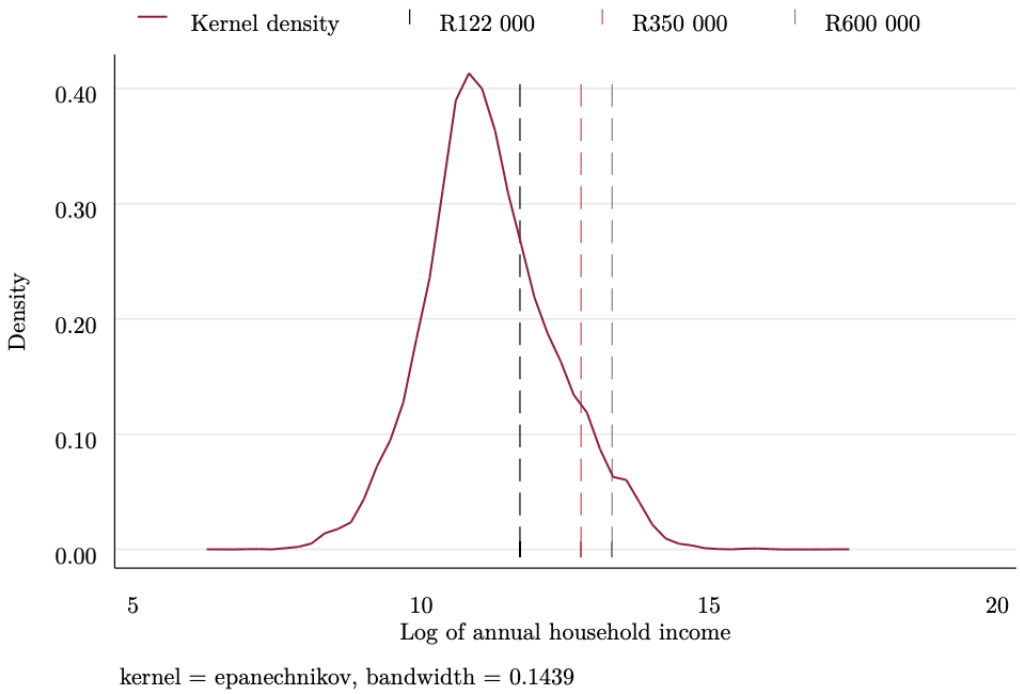
Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period and corrected for panel attrition.

2.7. Profiles by funding class

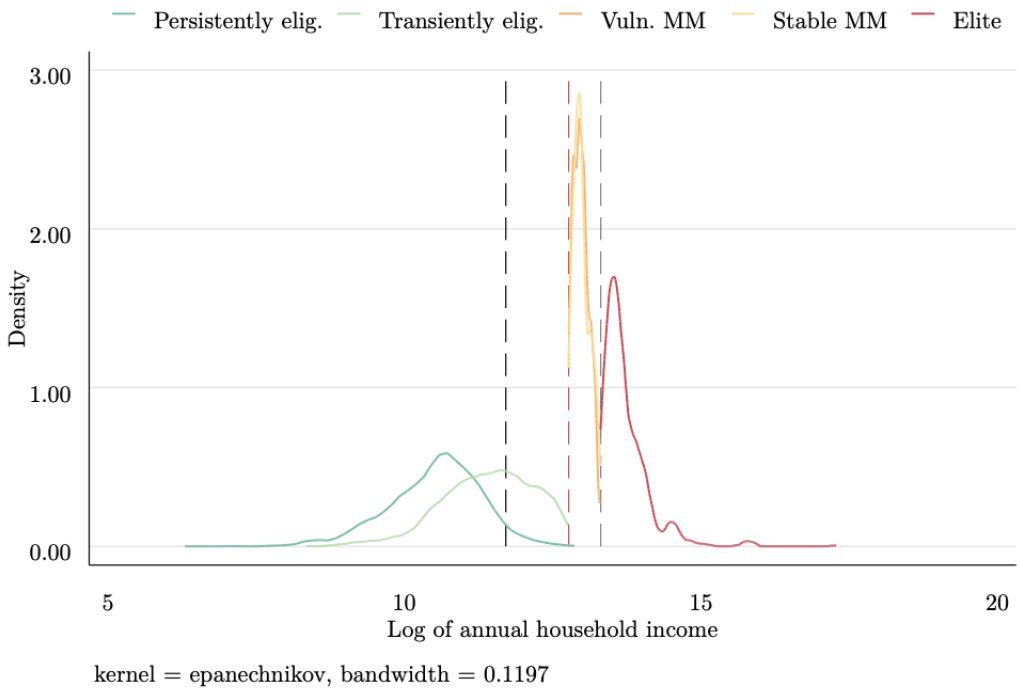
In Figure 2.4, I refine the classification of households shown in Figure 2.1 by taking economic (in)stability into account. The household income distributions of the persistently eligible and the transiently eligible are distinct, indicating that the transiently eligible are better off, on average, in terms of household income. On the other hand, the income distributions of the vulnerable and stable missing middle groups are very similar. As I describe in Section 2.7.2, there is nonetheless compelling variation in household circumstances among the missing middle that speaks to relative economic vulnerability and stability.

Figure 2.4 Distribution of household income in South Africa, 2017 (updated to reflect funding class)

(a) Annual household income (log) - policy stratification



(b) Annual household income (log) - our stratification schema



Source: Author's own calculations using NIDS Wave 5 (post-stratified weights).
Notes: Distribution in levels is trimmed at the 99th percentile.
One observation per household, income in December 2017 Rands.

2.7.1. Household circumstances of the NSFAS eligible

Table 2.7 shows that over the period 2008 to 2017, approximately 2% of the population fell within the vulnerable missing middle, with a marginally larger share classified as the stable missing middle. As previously noted, a large majority of South Africans have household income below the R350 000 threshold, with 60% of the population facing below average probabilities of moving above it (i.e. they are persistently eligible in my terminology). These households tend to be disproportionately headed by women who are unemployed, African, and who have low levels of completed education (primary or incomplete secondary). Households in the persistently eligible group have an equal reliance on governments grants and labour market income (39% of household income) on average, and access to credit markets and assets is low. The average per capita expenditure of those in the persistently eligible group is just R877.44, situating the average individual in this group below the upper bound poverty line.[39]

Those who are transiently eligible are clearly distinct from those persistently eligible along these above-mentioned dimensions. For example, over half the individuals in this group reside with a household head who is employed (59% compared to 37% among the persistently eligible) and a much larger share reside with a household head who has a post-school qualification (38% compared to 3% among the persistently eligible). Reliance on government grants diminishes among this group (15% vs. 39%) and over two-thirds of household income comes from the labour market on average, possibly reflecting the higher household employment rate for this group (55% of working age residents employed vs. 32% among the persistently eligible). Moreover, access to credit markets appears markedly higher for those transiently eligible compared to those persistently eligible.

Some similarity between the transiently eligible and the vulnerable missing middle is to be expected, since these groups are those whose household income is most likely to churn around the funding threshold. For example, the household employment rate is comparable across these groups, as is the share of individuals living with female household heads and household heads who have a post-school qualification. However, Table 1 makes it apparent that those who are transiently eligible are, in some respects, distinct from the vulnerable missing middle group. More than double the share in the vulnerable missing middle live with someone who has a home loan or bond compared to those who are transiently eligible (9% vs. 24%) and over three times the share in the vulnerable missing middle live with someone who has accessed vehicle finance (8% vs. 29%). In a sense, these results suggests that the current policy is relatively well targeted to the most resource-constrained households.

[39] R1 158 in 2017 (December 2017 prices).

Table 2.7 Average characteristics of households and household heads by funding class

	Eligible		Missing Middle			
	Persistent	Transient	Vulnerable	Stable	Elite	Total
Share of respondents	60.56%	31.09%	2.17%	2.41%	3.76%	100%
(a) Characteristics of the head						
Household head is employed	0.37	0.59	0.64	0.79	0.78	0.47
Age of household head	47.43	45.4	49.1	44.69	46.64	46.74
Female household head	0.64	0.47	0.48	0.26	0.28	0.56
Race						
African	0.92	0.72	0.76	0.16	0.24	0.81
Coloured	0.06	0.13	0.13	0.13	0.08	0.09
Asian/Indian	0.01	0.03	0.05	0.06	0.1	0.02
White	0.00	0.12	0.06	0.65	0.58	0.08
Highest education						
Missing	0.00	0.01	0.00	0.01	0.01	0.00
No schooling	0.20	0.02	0.03	0.01	0.00	0.13
Primary	0.31	0.13	0.09	0.00	0.03	0.23
Incomplete secondary	0.38	0.24	0.29	0.08	0.09	0.32
Matric	0.07	0.21	0.22	0.19	0.15	0.12
Post-school qualification	0.03	0.38	0.37	0.71	0.72	0.19
(b) Household characteristics						
Income and expenditure						
Per capita expenditure[a]	877.04	2 872.53	5 265.14	9 943.06	14 903.00	2 339.01
Per capita income[a]	1 209.42	3 279.34	8 835.43	11 986.31	31 796.22	3 429.11
Share of income from[b]:						
Labour market	0.39	0.70	0.80	0.88	0.83	0.53
Government grants	0.39	0.14	0.05	0.01	0.01	0.28
Investment income	0.01	0.03	0.03	0.04	0.07	0.02
Remittances	0.07	0.04	0.03	0.01	0.01	0.06
Subsistence agriculture	0.01	0.00	0.00	0.00	0.00	0.00
One-shot response (no source)	0.12	0.09	0.08	0.05	0.07	0.10
Assets, credit, and infrastructure						
At least one resident has a:						
Home loan/bond	0.00	0.09	0.24	0.53	0.47	0.07
Bank loan	0.07	0.22	0.34	0.28	0.22	0.13
Study loan with a bank	0.00	0.01	0.03	0.03	0.03	0.01
Vehicle finance	0.00	0.09	0.30	0.35	0.39	0.06
Credit card	0.02	0.14	0.32	0.47	0.51	0.09
Store card	0.18	0.42	0.55	0.44	0.42	0.28
Financial assets	0.59	0.90	0.94	0.94	0.95	0.71
Pension assets	0.01	0.17	0.31	0.52	0.47	0.09
Household member owns dwelling	0.76	0.71	0.86	0.78	0.85	0.75
Household owns livestock assets	0.09	0.02	0.04	0.00	0.01	0.06
Number of rooms in house	3.78	4.95	5.47	5.93	6.78	4.34
Access to electricity	0.79	0.95	0.94	0.97	0.98	0.85
Piped water on site	0.64	0.92	0.94	0.99	0.96	0.75
Has a flush toilet	0.40	0.81	0.82	1.00	0.97	0.57
Household composition						
Number of household residents	5.47	5.39	5.84	3.78	4.03	5.36
Number of children under 6	0.96	0.69	0.83	0.33	0.37	0.84
Number of children aged 6–18	1.74	1.58	1.56	0.87	0.96	1.63
Number of elderly residents[c]	0.34	0.30	0.28	0.26	0.25	0.32
Household employment rate[d]	0.32	0.55	0.57	0.74	0.70	0.42
Location						
Traditional	0.46	0.20	0.19	0.04	0.05	0.35
Urban	0.47	0.77	0.78	0.95	0.94	0.60
Farm	0.06	0.02	0.03	0.01	0.02	0.05

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period.
[a] Monthly value in Rands. [b] Share excludes imputed rental income. Labour market income reflects after tax income.
[c] Aged 60+. [d] Of working age adults.

A key priority of this work is to locate youth (those aged 15-35) within this framework. Education and training is a promising way of connecting NEETs to the labour market, making a focus on youth circumstances in relation to student funding particularly relevant in the South African context. Table A.3 in Appendix A replicates the characteristics shown in Table 2.7, but for the sample of youth rather than all respondents.[40] The general observations made here hold true among youth too. In the following section, I spotlight the differences between the vulnerable and stable missing middle groups among youth (aged 15-35) and show how grouping all youth in the missing middle together could mask important variation in their household circumstances.

2.7.2. Household circumstances of youth in the missing middle

Of the key differences between vulnerable and stable missing middle youth visible in Table 2.8, many may be attributed to the urban-rural split between groups.[41] While 76% of youth in vulnerable missing middle reside in urban locations, a much larger share of stable missing middle youth reside in urban areas (94%). On average, stable youth live in households that have better access to the labour market (73% household employment rate among the stable compared to 54% among the vulnerable), resulting in a greater share of household income from the labour market – ten percentage points more than the vulnerable. Furthermore, over double the share of stable missing middle youth live in a household with access to a home loan/bond, compared to youth who are in the vulnerable missing middle (54% vs. 21%), and a much lower share of vulnerable missing middle youth reside in households in which a resident has access to vehicle finance and pension assets. Stable missing middle youth also reside in households that are smaller on average, have fewer dependants, and are more likely to be headed by an employed male with a post-school qualification. The fact that the stable missing middle have such a high share of household heads with post-school education is relevant since first-hand experience of application and enrolment could facilitate access.

Comparing the averages for both of these groups individually to those in the final column, it is apparent that aggregating household circumstances within the missing middle would overstate the long-run well-being of the vulnerable group and understate the relative economic security of the stable group. For example, it would appear as if the share of the missing middle with access to a home loan/bond in their household is just over a third, when in fact it is over half for the stable group but below a quarter among the vulnerable.

This being said, household circumstances are not the only factors affecting access to post-schooling. A number of individual characteristics such as schooling, academic performance, and family background are also relevant, and it is useful to understand how these factors vary by funding class. Table 2.9 summarises this information.

[40] A comparison of youth households vs. all households is shown in Table A.4 in Appendix A.
[41] See Figure A.2 in Appendix A for the distribution of youth by province and funding class.

Table 2.8 Average household characteristics of youth (aged 15-35) in the missing middle

	Vulnerable	Stable	P-value*	All MM
(a) Characteristics of the head				
Household head is employed	0.65	0.84	0.00	0.73
Age of household head	49.13	40.95	0.00	44.69
Household head is female	0.46	0.26	0.00	0.37
Race				
African	0.79	0.21	0.00	0.48
Coloured	0.13	0.13	0.85	0.12
Asian/Indian	0.06	0.08	0.47	0.07
White	0.02	0.58	0.00	0.32
Highest education				
Missing	0.00	0.01	0.02	0.00
No schooling	0.03	0.01	0.01	0.02
Primary	0.11	0.00	0.00	0.05
Incomplete secondary	0.33	0.05	0.00	0.19
Matric	0.20	0.17	0.47	0.19
Post-school qualification	0.35	0.76	0.00	0.55
(b) Household characteristics				
Income and expenditure				
Per capita expenditure[a]	5 026.34	10 163.72	0.00	7 664.28
Per capita income[a]	8 748.65	12 609.23	0.00	10 705.20
Share of income from[b]:				
Labour market	0.81	0.91	0.00	0.85
Government grant	0.05	0.01	0.00	0.02
Investment income	0.03	0.03	0.96	0.03
Remittances	0.04	0.02	0.07	0.03
Subsistence agriculture	0.00	0.00	0.02	0.00
One-shot response (no source)	0.08	0.04	0.03	0.06
Assets, credit, and infrastructure				
At least one resident has a:				
Home loan/bond	0.21	0.54	0.00	0.38
Bank loan	0.35	0.30	0.20	0.32
Study loan with a bank	0.05	0.04	0.47	0.05
Vehicle finance	0.28	0.40	0.00	0.35
Credit card	0.31	0.49	0.00	0.41
Store card	0.56	0.46	0.02	0.52
Financial assets	0.94	0.94	0.98	0.94
Pension assets	0.32	0.49	0.00	0.40
Household member owns dwelling	0.83	0.73	0.02	0.77
Household livestock assets	0.03	0.00	0.00	0.02
Number of rooms in house	5.58	5.84	0.16	5.74
Household has access to electricity	0.95	0.98	0.03	0.97
Piped water on site	0.93	0.99	0.00	0.96
Has a flush toilet	0.81	1.00	0.00	0.90
Household composition				
Number of household residents	6.04	3.70	0.00	4.81
Number of children under 6	0.84	0.35	0.00	0.58
Number of children aged 6–18	1.44	0.70	0.00	1.06
Number of elderly residents[c]	0.28	0.16	0.01	0.22
Household employment rate[d]	0.54	0.73	0.00	0.64
Location				
Traditional	0.21	0.04	0.00	0.12
Urban	0.76	0.94	0.00	0.86
Farm	0.03	0.02	0.60	0.02
No. individuals	929	517	-	1 515

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: * P-value for the difference between vulnerable and not vulnerable mean characteristics. Data are weighted using post-stratification weights from the base period.
[a] Monthly value in Rands. [b] Share excludes imputed rental income. Labour market income reflects after tax income.
[c] Aged 60+. [d] Of working age adults.

Table 2.9 Average individual characteristics of youth (aged 15-35) by funding class

	Eligible		Missing Middle			
	Persistent	Transient	Vulnerable	Stable	Elite	Total
Activity in time t-1						
Enrolled but level is unknown	0.00	0.01	0.01	0.01	0.02	0.01
Enrolled in school	0.25	0.19	0.24	0.24	0.23	0.23
Enrolled in post-school	0.04	0.09	0.15	0.14	0.22	0.06
Employed	0.28	0.45	0.36	0.53	0.44	0.34
NEET[a]	0.43	0.26	0.23	0.09	0.08	0.36
Community unemployment rate	0.65	0.62	0.63	0.60	0.59	0.64
Demographics						
Age	24.18	25.17	24.17	25.07	24.33	24.52
Gender (Male)	0.48	0.52	0.49	0.47	0.46	0.49
Race						
African	0.93	0.76	0.79	0.22	0.30	0.84
Coloured	0.06	0.12	0.12	0.13	0.09	0.08
Asian/Indian	0.01	0.03	0.06	0.08	0.11	0.02
White	0.00	0.09	0.02	0.57	0.49	0.06
Schooling						
School quintile 1–3	0.75	0.54	0.43	0.08	0.16	0.64
School quintile 4–5	0.15	0.30	0.34	0.62	0.53	0.22
School quintile missing, with match[b]	0.02	0.03	0.06	0.05	0.06	0.03
School quintile missing, no match[b]	0.09	0.13	0.17	0.24	0.25	0.11
Academic performance						
Completed or enrolled in Grade 9	0.82	0.94	0.96	0.97	0.99	0.87
Completed or enrolled in Grade 12	0.29	0.56	0.59	0.76	0.75	0.40
Repeated a school grade[b]	0.56	0.47	0.38	0.19	0.22	0.51
Repeated a school grade missing	0.14	0.14	0.18	0.22	0.17	0.14
Repeated a secondary grade	0.28	0.24	0.17	0.06	0.08	0.25
Repeated a secondary grade missing	0.34	0.38	0.41	0.44	0.49	0.36
Repeated a primary grade	0.19	0.12	0.09	0.09	0.06	0.16
Repeated a primary grade missing	0.34	0.38	0.42	0.44	0.49	0.36
Standardised numeracy score[c]	-0.64	-0.44	-0.14	0.08	-0.08	-0.55
Mother's education[b]						
Missing	0.05	0.05	0.05	0.03	0.05	0.05
No schooling	0.25	0.11	0.09	0.03	0.02	0.19
Primary	0.32	0.20	0.12	0.05	0.08	0.27
Incomplete secondary	0.28	0.31	0.31	0.18	0.11	0.28
Matric	0.05	0.13	0.09	0.24	0.20	0.08
Post-school qualification	0.04	0.20	0.33	0.46	0.53	0.12
Father's education[b]						
Missing	0.18	0.16	0.16	0.05	0.10	0.17
No schooling	0.30	0.14	0.12	0.04	0.04	0.23
Primary	0.22	0.16	0.09	0.04	0.03	0.19
Incomplete secondary	0.19	0.25	0.22	0.13	0.11	0.20
Matric	0.07	0.14	0.19	0.30	0.20	0.11
Post-school qualification	0.04	0.15	0.22	0.44	0.52	0.10
No. individuals	38 245	15 654	929	517	745	56 090

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period.
[a] This category includes those with missing enrolment and employment information.
[b] Parental education, grade repetition, and school quintile are all constructed using the full panel. That is, missing gaps are filled in with data from previous waves, when available. We include indicator variables for missing information where we are unable to fill gaps with the panel data. For school quintile, we capture two binary variables depending on the type of missing information. School quintile may be missing because a private school was attended. Alternatively, it may be missing because the name of the school could not be matched to a master list of schools. We thus capture whether information is missing but a match in school names did occur (potentially private school enrolees), or whether it is missing and no match occurred. Grade repetition was asked about in Waves 1 and 5 only, to individuals aged 30 and under. Since our sample is aged 15-35, older members are more likely to have missing grade repetition information. Information from enrolment in between these two waves can be used to piece together grade repetition information in every wave.
[c] This variable was only asked in Wave 1.

The highest share of NEET youth (43%) are found among those who face the most economic precarity (the persistently eligible), perhaps a reflection of poorer schooling outcomes among this group. This is compared to a low 8% and 9% of youth who are NEET among the elite and stable groups, respectively. Roughly a quarter of youth in each group are still enrolled in school.

The age and gender distribution across groups is similar, but notable differences appear in school quintile and academic performance. The majority of youth whose household income is below the R350 000 threshold attended a quintile 1-3 school (75% and 54% for the persistently and transiently eligible respectively), whereas attending a quintile 4-5 school is more prevalent among the stable missing middle (62%). The vulnerable group falls somewhat in-between the transiently eligible and stable, but more vulnerable youth attend(ed) a quintile 1-3 school (43%) than a quintile 4-5 school (34%). Here, youth in the vulnerable missing middle appear more similar to transiently eligible youth, on average, than the stable missing middle.

Higher incidences of grade repetition and lower numeracy scores are prevalent among those persistently and transiently eligible, a stark reminder that household socioeconomic status and schooling outcomes are inextricably linked. Moreover, less than one third of youth who are persistently eligible reach Grade 12, despite their average age being the same as youth in the vulnerable missing middle. Over half of persistently eligible youth have repeated at least one school grade, and grade repetition remains high among vulnerable missing middle youth – over one third have completed a school grade – this being double the share who ever repeated among the stable group. The driving force appears to be repetition in secondary school grades.

Family background as captured by parental education is also distinct for the vulnerable vs. stable missing middle youth. Double the share of stable youth have a father with post-schooling compared to vulnerable youth (44% vs. 22%), and more than double the share have a mother with a matric (24% vs. 9%). Lastly, it is impossible to ignore the nature of inequality along racial lines. Although both the vulnerable and stable collectively comprise the missing middle, the vulnerable missing middle are disproportionately African (79%) whereas the stable missing middle are disproportionately white (57%).

These results emphasise that by solely using a monetary threshold to classify individuals, there is a risk of mis-identifying a share of the population whose household income puts them in the missing middle but is nevertheless relatively economically secure. These individuals live in smaller households with fewer dependents (both children and elderly) that are more connected to the labour market and have relatively greater access to loans. In fact, in many respects, the stable middle class more closely resembles the elite, on average, compared to the vulnerable missing middle. I highlight these findings not to suggest that those in the missing middle do not face barriers to access but rather to prompt a reflection about differentiated targeting of policy.

2.8. Caveats and considerations

2.8.1. Sensitivity to the choice of probability cut-off

I recognise that the size and mean characteristics of my funding classes may be sensitive to the value set as the probability threshold. In the main analysis, I follow Schotte et al. (2018) and López-Calva and Ortiz-Juarez (2014) in setting the cut-off separating the vulnerable and stable missing middle at the observed rate of those with household income above the funding threshold that fell below in the next period (37.79%). That is, those with predicted probabilities of household income falling below the threshold of more than 37.79% are classified as vulnerable.

In principle, however, it would be possible to make an argument for the use of another probability threshold that sets a different cut-off for the vulnerable. By way of illustration, a stricter (higher) threshold could be implemented as the cut-off if, for example, scarcity of funds is a concern. In this hypothetical case, I am interested in the extent to which this would change who is defined as vulnerable. A concern would be that through the implementation of a stricter threshold, particularly vulnerable individuals will be classified as not requiring support.

Key to understanding the sensitivity of my profiling to these thresholds, is to understand how the estimates of mean household characteristics change when the missing middle is split by different probability thresholds. To explore this, I examine the average differences in key characteristics of the vulnerable and stable missing middle groups when the threshold separating these groups is raised from 37.79% to 50% and 75%, respectively. The vulnerable missing middle will then be defined as those whose predicted probabilities of their household income falling below R350 000 are more than 0.5 and more than 0.75, respectively. Intuitively, by raising the threshold the share of the missing middle classified as vulnerable will be lower, and the share classified as stable will be larger. That is, as the threshold value increases, those whose predicted probabilities of household income between 37.79% and 75% will be reclassified from vulnerable to stable.[42] Table 2.10 shows these results.[43]

When defined by an average probability of falling below R350 000 being is greater than 0.5, 35% of the missing middle is classified as vulnerable. When the vulnerable group is restricted to those whose average probability of falling below R350 000 is above 0.75, only 16% of the missing middle remain classified as vulnerable. I examine the mean characteristics of the vulnerable and stable missing middle groups under these compositional changes. A couple of main points stand out. First, as individuals move from vulnerable to stable groups, it is apparent that although mean characteristics of both groups fall, the stable missing middle do not experience as large a

[42] Throughout this section, I consider which individuals in the missing middle are classified as vulnerable and stable when setting different values as the probability threshold. However, the same logic applies to the threshold separating the persistently eligible from the transiently eligible.

[43] Table A.5 in Appendix A shows the average of all characteristics of the household and household head for the vulnerable and stable missing middle when the probability threshold is set to 0.5 and 0.75, respectively.

drop in their mean characteristics as the vulnerable. This is most evident in income per capita, where one observes a greater fall in mean per capita income when relatively less vulnerable individuals are pushed into the stable class, compared to the fall in mean per capita income for the stable group when they absorb some individuals moving up from the vulnerable group.

Table 2.10 Average characteristics of households and household heads for those in the vulnerable vs. stable missing middle by probability threshold value

	Threshold = 37.79%		**Threshold = 50%**		**Threshold = 75%**	
	Vulnerable	Stable	Vulnerable	Stable	Vulnerable	Stable
Share of missing middle	47.40%	52.60%	34.52%	65.48%	15.66%	84.34%
(a) Characteristics of the head						
Race						
African	0.76	0.16	0.83	0.24	0.94	0.35
Coloured	0.13	0.13	0.11	0.14	0.03	0.15
Asian/Indian	0.05	0.06	0.04	0.07	0.03	0.06
White	0.06	0.65	0.02	0.56	0.00	0.44
Household head is employed	0.64	0.79	0.65	0.75	0.52	0.75
Female household head	0.48	0.26	0.51	0.29	0.59	0.32
Head has a post-school qualification	0.37	0.71	0.32	0.67	0.15	0.62
(b) Household characteristics						
Household annual income	446 873.20	456 642.00	443 649.80	456 420.10	440 593.50	454 131.30
Per capita monthly income[a]	8 835.43	11 986.31	8 218.88	11 691.68	6 233.05	11 283.60
Share of income from labour market	0.80	0.88	0.79	0.88	0.69	0.88
At least one resident has:						
Financial assets	0.94	0.94	0.93	0.94	0.92	0.94
Pension assets	0.31	0.52	0.29	0.48	0.24	0.45
Home loan/bond	0.24	0.53	0.18	0.51	0.09	0.46
Bank loan	0.34	0.28	0.32	0.31	0.29	0.32
Vehicle finance	0.30	0.35	0.27	0.35	0.16	0.36
Credit card	0.32	0.47	0.31	0.45	0.23	0.43
Household dependency ratio	0.63	0.49	0.64	0.51	0.79	0.51
Household employment rate[b]	0.57	0.74	0.54	0.72	0.48	0.69
Household is in an urban location	0.78	0.95	0.75	0.93	0.66	0.90

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period.
[a] Value in December 2017 Rands. Labour market income reflects after tax income. Share from labour market reflects the share of household income from the labour market. [b] Of working age adults.

Second, when only 16% of the missing middle remains in the vulnerable group (those whose average probability of falling below the threshold is greater than 0.75), their average characteristics very closely reflect those of the transiently eligible group below the threshold. If one considers vulnerability from a mobility perspective, it is thus clear that these individuals in particular are warranting of policy attention. This result further speaks to the potential benefits of a more nuanced or differentiated funding instrument, especially for those in the missing middle. Indeed, it may be more tangible and palatable to extend funding to the most vulnerable 16% of the missing middle, rather than the full spectrum of incomes off which it is currently defined.

Income thresholds are always likely to exist for operational purposes and thus speculating about an increase in a *probability* threshold is purely hypothetical. Nonetheless, it is reassuring that the above results suggest that by raising the threshold there does not appear to be a considerable risk of mis-identifying highly vulnerable individuals as 'stable' – at least not in ways that substantially change average access to credit markets, assets, and household employment. This could be because the current household income threshold of R350 000 already does a relatively good job of capturing household vulnerability.

2.8.2. The assumption underlying pooled transitions

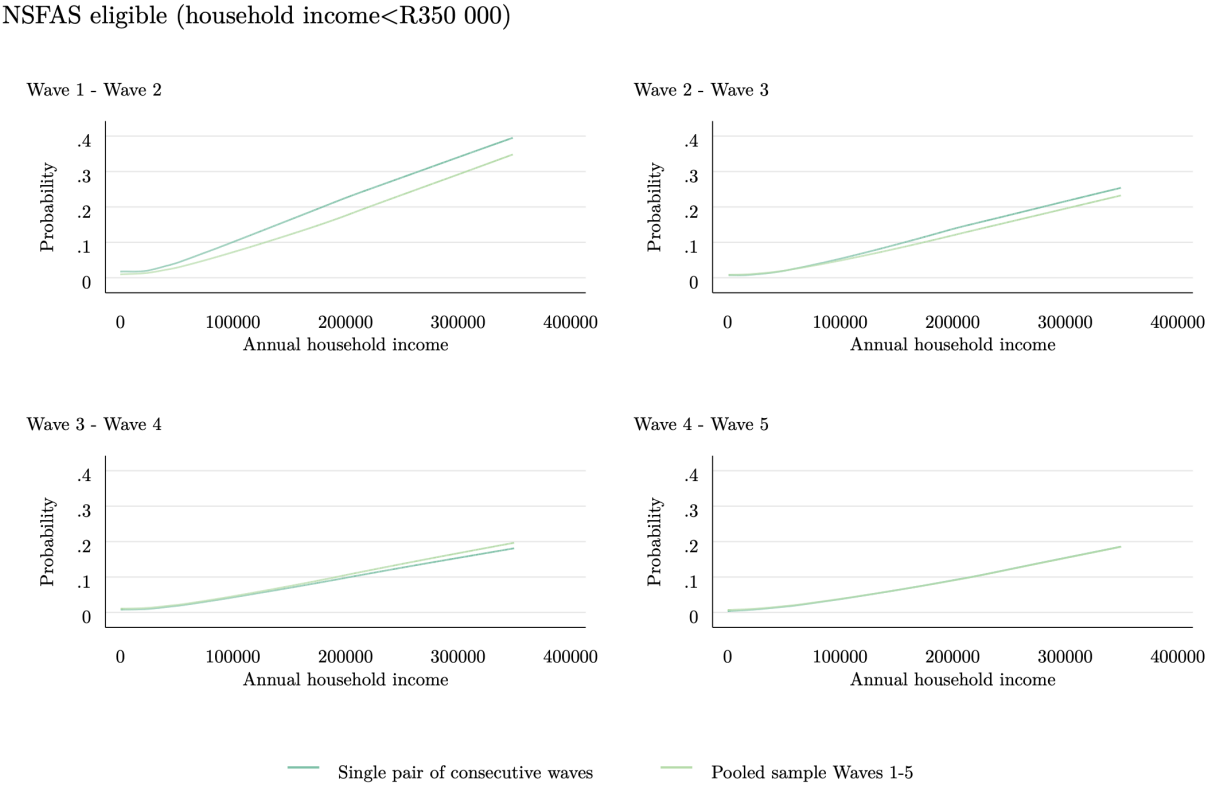
By pooling waves of observations to assess vulnerability, I implicitly assume that the economic conditions affecting transitions in each wave remain unchanged throughout the panel. As noted by Zizzamia et al. (2016), this is an important assumption given that changes in the macroeconomic environment will affect the chances of household income rising or falling. In order to examine whether the process generating a rise or fall in household income is reasonably consistent between waves, I fit the multivariate probit regression to each pair of waves independently, full results of which are presented in Table A.6 in Appendix A.[44] A majority of covariates remain consistent in terms of sign and significance. In most of the cases where the sign of the effect differs, the effect of the covariate is not statistically different from zero. This ameliorates some concerns about changing environments between waves.

Nonetheless, it is the predicted conditional propensities of household income moving above or falling below the threshold that are key to my analysis, rather than the coefficient estimates. Figures 2.5 and 2.6 therefore compare the distribution of predicted propensities for the eligible and missing middle groups when predicted from the multivariate probit on the pooled sample and each pair of waves, respectively.

The fact that the distributions from the pooled sample and each pair of waves track each other closely reassures me that I am not misidentifying funding classes by classifying them off predicted conditional propensities using the pooled sample rather than separate wave by wave transitions. However, the distributions in the figures do draw attention to the comparability of predicted propensities between pairs of waves, particularly between the Wave 1-2 transition and the Wave 2-3 transition. For the Wave 2-3 transition in Figure 2.5, at each income level the average predicted probability of household income moving above the funding threshold is lower compared to the Wave 1-2 transition. This may owe to the fact that in Wave 1, the sample of respondents is nationally representative, but in each subsequent wave, attrition affects the extent to which the sample (and thus the next 'base period') remains representative. Therefore, those

[44]Note that I do not include interactions between initial state and the covariates in these specifications, since the model does not converge when interactions are included. The probit specification without interactions mirrors that used by López-Calva and Ortiz-Juarez (2014) for identifying a vulnerable middle class.

Figure 2.5 Distribution of predicted conditional probabilities of household income moving above the funding threshold



Source: Author's own calculations using NIDS Waves 1-5 pooled sample.
Notes: Annual income in December 2017 Rands.
Lowess smoothing of the predicted conditional probabilities for household income moving above R350 000.

who would be more likely to transition above the threshold between Waves 2 and 3 were more likely to be lost from the sample between Waves 1 and 2. By design of the sample, I could thus be underestimating the size of the transiently eligible group. On the other hand, one could argue that the size of the transiently eligible group is over-estimated. This would be that case if we are disproportionately losing people who would have moved above the threshold.

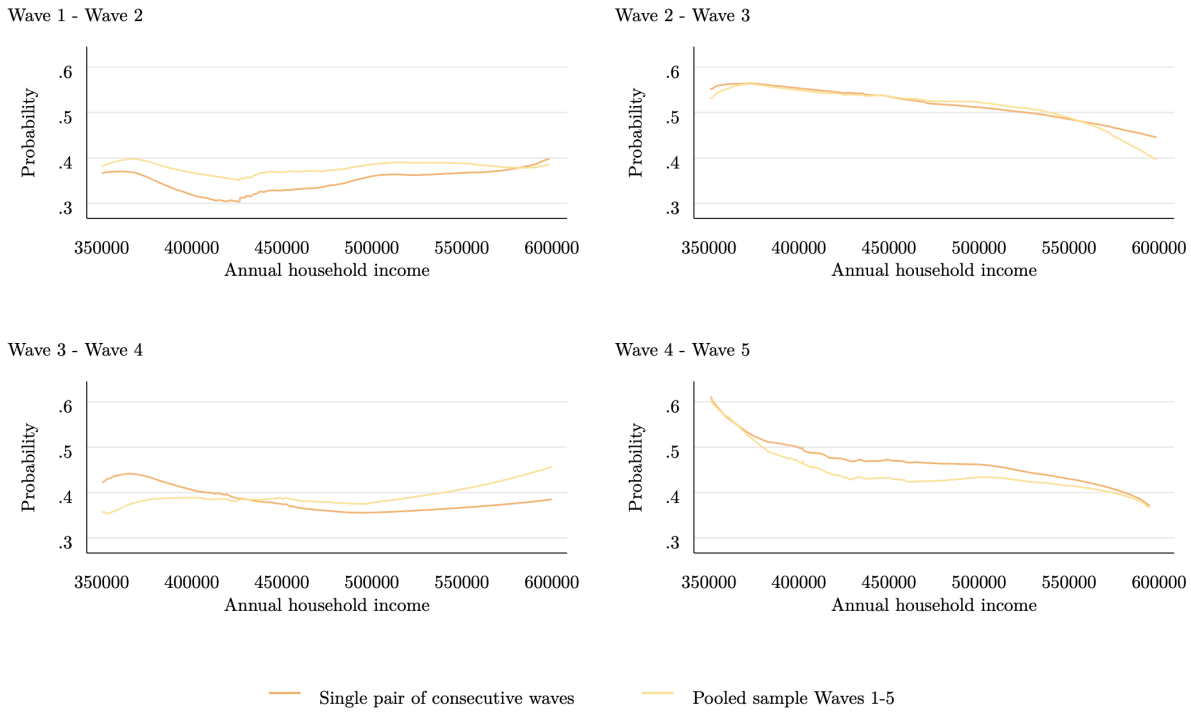
Although the post-stratification weights aim to ensure that each cross section is representative of the population at the time, expecting the weights to be able to fully account for information lost due to attrition is, perhaps, too ambitious.

In Figure 2.6, I observe that for the Wave 2-3 transition predicted probabilities of falling below the funding threshold are higher at each income level compared to the Wave 1-2 transition. This is likely due to the fact that those whose household income is more likely to remain above the threshold (and thus have lower predicted propensities of falling below) would have been more likely to attrit between Waves 1 and 2. Base period Wave 2 thus excludes these individuals. The difference in predicted probabilities is less stark when comparing the predictions from the pooled sample, however. Nonetheless, I likely underestimate the size of the stable missing middle.

A further note on attrition is relevant here. Although the multivariate probit regression

Figure 2.6 Distribution of predicted conditional probabilities of household income falling below the funding threshold

Missing middle (household income R350 000 - R600 000)



Source: Author's own calculations using NIDS Waves 1-5 pooled sample.
Notes: Annual income in December 2017 Rands.
Lowess smoothing of the predicted conditional probabilities for household income moving above R350 000.

accounts for attrition directly, it is important to be aware that I include TSMs in my base period sample. This ensures that I have a nationally representative sample, or as close to that as possible, when using post-stratification weights in the base period. However, unless TSMs remain co-resident with a CSM in the following wave, they are not actually tracked. Therefore, the inclusion of TSMs in the base period sample may misrepresent the type of attrition process from base period waves in which TSMs are present (i.e. Waves 2-4),[45] since TSMs will drop out of the study the moment that they cease to co-reside with a CSM. Although I include a binary variable in the retention equation for whether a respondent is a CSM or TSM, this only partly addresses this concern. The fact that I correct for both types of attrition in the pooled sample, but no TSMs drop out between Waves 1 and 2, may be why I observe better aligned distributions in the transitions between Waves 2-3, Waves 3-4, and Waves 4-5 compared to the jump in distributions between the Waves 1-2 and Waves 2-3 transitions.

Whether the size of certain groups is over- or under-estimated is thus difficult to disentangle, and concerns about attrition are not easily resolved, even with weights. As a robustness check

[45] There are no TSMs in Wave 1.

for my funding classes defined from the pooled sample, I reclassify my strata based on the Wave 1-2 transition only, since Wave 1 is the most nationally representative base period, by design. In this case, the probability threshold between the vulnerable and stable missing middle is set at 37.20% – the observed rate of household income falling below the threshold in between Wave 1 and Wave 2.

Table 2.11 compares key households characteristics for the vulnerable and stable missing middle when these groups are estimated using the pooled sample (as in the main analysis) and the Wave 1-2 transition only.[46]

Table 2.11 Average characteristics of households and household heads in the missing middle (pooled transition vs. wave 1-2 transition)

	Vulnerable missing middle		Stable missing middle	
	Pooled transitions	Wave 1-2 transition	Pooled transitions	Wave 1-2 transition
Weighted share of respondents	2.17%	1.94%	2.41%	1.73%
(a) Characteristics of the head				
Race				
African	0.76	0.70	0.16	0.06
Coloured	0.13	0.12	0.13	0.10
Asian/Indian	0.05	0.12	0.06	0.04
White	0.06	0.06	0.65	0.79
Female household head	0.48	0.22	0.26	0.14
Household head is employed	0.64	0.74	0.79	0.76
Household head has a post-school qualification	0.37	0.43	0.71	0.54
(b) Household characteristics				
Per capita monthly income[a]	8 835.43	9 379.95	11 986.31	11 121.68
Share of household income from labour market	0.80	0.85	0.88	0.95
At least one resident has:				
Financial assets	0.94	0.99	0.94	0.99
Pension assets	0.31	0.27	0.52	0.51
Home loan/bond	0.24	0.29	0.53	0.66
Bank loan	0.34	0.38	0.28	0.26
Vehicle finance	0.30	0.47	0.35	0.34
Credit card	0.32	0.45	0.47	0.61
Household dependency ratio	0.63	0.63	0.49	0.51
Household employment rate[b]	0.57	0.55	0.74	0.66
Household is in an urban location	0.78	0.89	0.95	0.97

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period.
[a] Value in December 2017 Rands. Labour market income reflects after tax income. [b] Of working age adults.

In general, the estimates of access to credit markets and financial assets are lower for both

[46] Only missing middle groups have been compared here, but a similar comparison can be made below the funding threshold. Table A.7 in Appendix A shows the average of all characteristics of the household and household head for all funding classes as defined from the Wave 1-2 transition.

the stable and vulnerable missing middle when using the pooled sample rather than the Wave 1-2 transition only. This may owe to the fact that individuals who access these markets are more likely to fall out of the sample over time. Similarly, the estimated share of household income from the labour market is lower from the pooled sample for both groups. The share with access to pension assets in the household, however, is higher when using the pooled sample rather than the Wave 1-2 transition only. Despite these differences in estimated shares, the practical implications of the findings in the main analysis remain unchanged.

2.9. Concluding remarks

This chapter proposes an empirical and conceptual methodology for analysing the NSFAS eligible and missing middle groups based on income mobility and the notion of vulnerability. I fit a Markov model to nationally representative panel data from the National Income Dynamics Study, and using coefficient estimates from a multivariate probit regression, predict the conditional probabilities of household income moving above or falling below the NSFAS income eligibility threshold in the near future. I then classify individuals based on their vulnerability to their household income remaining below or falling below the funding threshold.

A key policy priority currently is to establish a sustainable, comprehensive, and progressive financial aid scheme to fund students enrolled in post-school education. If free education for all students is not viable, a differentiated policy might be necessary. A key value in my approach lies in operationalising the concepts of mobility, vulnerability, and economic stability to differentiate the socioeconomic circumstances of households within the context of the current post-school funding policy. My results showcase that the missing middle is a complex category, comprising two distinct groups based on their relative economic stability or vulnerability. In focussing on youth, I show that aggregating household circumstances within the missing middle would overstate the long-run well-being of vulnerable youth and understate the relative economic security of youth I classify as relatively more stable.

While income thresholds are always likely to exist for operational purposes in a social support environment, my work shows that when considering the design of support, 1) the current NSFAS funding threshold should not be lowered – it is well targeted at the those living in the most resource-constrained households, and 2) when considering extending support to the missing middle, a differentiated funding instrument could work. This group looks far more likely to be engaging in formalised credit markets, for example. That being said, I acknowledge that vulnerability to income fluctuations can be an important constraint to post-school access – even when income is above the threshold. Rising student debt is evidence alone that better household circumstances do not equate to an ability to afford post-school education.

Overall, this chapter has made two important conceptual and methodological contributions. First, conceptually, it establishes a well-defined framework for the analysis of the missing middle and other eligible groups as they relate to funding thresholds, over time. For example, future

research may consider the same conceptualisation of vulnerability as it applies to those who became newly eligible in 2018. That is, those who fall in the R122 000 to R350 000 income bracket. This group is also of policy relevance, given that the current funding model, in which they are included, is regarded as unsustainable. Second, from a methodological point of view, to the best of my knowledge this is the first study to apply an income dynamics approach to studying funding thresholds, while addressing simultaneously the endogeneity of initial conditions and panel attrition.

Chapter 3
Learning in lockdown: University students' academic performance during COVID-19 closures

Abstract

In response to the COVID-19 pandemic, South African higher education institutions closed and rapidly moved to implement remote learning solutions. Many students lacked access to data and learning devices, and structural inequalities shape the household environments to which many students returned – and in which they were expected to learn new academic material. To date, however, there has been a paucity of research quantifying the effects of the pandemic on learning loss and academic performance among South African university students.

As such, I explore changes in university students' academic performance during the pandemic using longitudinal, administrative data from the University of Cape Town. I estimate a difference-in-differences model in which I allow for the effect of the pandemic on students' outcomes to differ by their academic performance in previous years as well as their socioeconomic status. Results show that although performance improved on average in 2020, this was driven by performance gains at the bottom end of the grade point average distribution. Overall, performance gains made in 2020 are reversed in 2021, suggesting academic performance improvements in 2020 did not reflect true learning gains. Of particular concern is a widening achievement gap between NSFAS-funded students and students not funded by NSFAS in 2021, suggesting household inequalities are playing out in achievement inequalities to a greater extent since COVID-19.

3.1. Background, objectives, and contributions

In response to the COVID-19 pandemic, South African tertiary education institutions and student residences around the country abruptly closed in order to mitigate the spread of COVID-19. As such, universities were required to come up with remote teaching and learning solutions in a relatively short period of time. However, many students lack off-campus access to internet, data connectivity, and devices (Pillay et al., 2021; Whitelaw et al., 2020), and glaring structural inequalities plague a multitude of socioeconomic factors in South Africa. These factors shape the household environment in which many students found themselves, and in which they were expected to learn new academic material.

Both locally and globally, evidence has unfolded that indicates substantial learning losses (or lack of learning progress) due to COVID-19-related school closures (Ardington, 2021; Aucejo et al., 2020; Maldonado & Witte, 2020). Based on this evidence, learning loss could be anticipated across the university sector too. That is, a variety of factors impacting both students and educators are likely to have worked to decrease learning quantity and quality in both school and university settings: for example, less time spent on education due to home responsibilities, lack of motivation, stress symptoms, insufficient access to digital infrastructure at home, inadequate digital skills for remote learning and teaching, varying degrees of financial and non-financial familial support, and studying in environments not conducive to learning. Nevertheless, literature quantifying the effects of the pandemic on learning loss and student performance among South African university students in particular is currently lacking.

Despite expectations of learning loss and incomplete learning experiences in the university sector (Salmi, 2020), emerging evidence from South African universities suggests academic performance improved due to a number of simultaneously occurring and interacting factors. These factors include increased marking leniency, a change in the content taught, different assessment practices, increased cheating,[47] students adopting better learning strategies through online learning,[48] and learning at their own pace. Thus, while it may be expected that learning losses occurred, identifying the extent of loss is hampered by heterogenous behaviour before and during the pandemic on the part of both students and staff. This means that grades, graduation, and other academic performance measures become more imprecise measures of learning in pandemic years.

Nevertheless, there are a number of reasons that documenting the effects of the lockdown on students' academic performance – whether positive or negative – is important. For the purposes of this chapter, 'performance', refers to measures of achievement, such as grades or credits earned

[47] A survey of academic staff across South Africa's universities suggests that 73% of respondents raised concerns about academic misconduct (Nkosi, 2021). Globally, increased cheating and plagiarism in online assessments have been a concern (Salmi, 2020).

[48] This is more likely to be the case for older learners (e.g. university students) than learners of school going age.

(the latter being the measure of interest for this analysis). This differs from actual learning, which could decline despite performance improvements. If there were unanticipated increases in student performance, this may be accompanied by dips in performance in coming years. That is, spuriously improved or even constant academic performance during the pandemic could unintentionally impact academic performance in the long term if students proceed to higher levels without sufficient baseline knowledge and competencies (Dandara et al., 2017). This was one of the main concerns expressed by academics at the University of Cape Town [UCT] during 2021; in particular the extent to which students promoted into next level courses had actually met the necessary learning prerequisites (Lange, 2021). This not only has implications for students while at university, it also has bearing on the quality of grades and graduation as signals in the labour market – especially since the majority of assessments are not standardised, and thus can be tailored to content taught. Moreover, increases or decreases in performance may be more pronounced for certain groups of students than others, and understanding how this may impact existing achievement gaps and inequalities in the system is relevant.[49]

Using longitudinal, institutional data from UCT,[50] I estimate a difference-in-differences [DD] model in which I allow for the effect of the pandemic to differ by a student's academic performance in previous years as well as their socioeconomic status [SES]. For the purposes of this chapter, SES refers to an individuals' access to economic resources, such as income, education, social networks. It is proxied for by receipt of state financial aid, which is closely linked to income position but also educational and social disadvantage.

Moreover, I examine academic performance in 2021, considering that changes in academic performance during the pandemic could affect performance in subsequent years. I am particularly interested in differentiating the effect of COVID-19 from general time trends[51] and how COVID-19 impacted existing achievement gaps between students. I thus aim to answer the following questions:

1. What was the impact of remote learning and assessment on UCT students' credits passed and credits taken in 2020 relative to their peers in pre-pandemic years?

2. Were university students on financial aid disproportionately affected by pandemic-related changes in 2020, and were there differences along the academic performance distribution?

[49] Relatedly, 2020 was a key year for policy evaluation, since the cohort enrolling under the new NSFAS policy instituted in 2018 would have been graduating from three-year degrees. Thus understanding the impact of the pandemic on student performance is critical to understanding the impact on NSFAS-funding policy evaluations too.

[50] The UCT data is substantially richer in the information that it captures about students compared to the corresponding data in the national database. Some of the information lacking in the national data is central to my empirical strategy.

[51] E.g. if student performance has been improving over time and the full performance improvement between 2019 and 2020 is attributed to COVID-19, the effect of the pandemic may be overestimated.

3. To what extent do students' credits taken and passed in 2021 differ from pre-pandemic trends?

The remainder of the chapter is organised as follows. In the following section, I review existing literature on the effects of institution closures on learning and student performance. Although literature on school closures is not unique to the COVID-19 pandemic (e.g. Allington et al., 2010), as Ardington (2021) notes, closures during COVID-19 were unprecedented in many ways: there was simultaneously occurring increased economic uncertainty, falling household income, rising unemployment, and increased psychological costs. I thus focus on pandemic-related closures in my review. Next, in Section 3.3, I provide institutional context and discuss UCT's response to COVID-19.[52] Thereafter, I provide the conceptual framework that guides my analytical approach in Section 3.4. I discuss my data and empirical strategy in Section 3.5, before presenting and discussing the results in Section 3.6. Finally, I conclude in Section 3.7.

3.2. Performance gains, learning losses, and contributing challenges

A growing number of studies on learning loss have emerged following the wake of COVID-19-related school closures globally. A key finding from this literature is that the extent of learning loss is distributed differently along gradients of academic proficiency and family SES. In South Africa, Ardington et al. (2021) use a DD approach to study learning losses in early grade reading, finding losses are most pronounced among the least proficient in reading, with more muted effects among those with higher initial proficiency. Ardington (2021) raises the concern that those who are behind may continue to learn less each year as they progress through their education. In this regard, Gustafsson and Nuga Deliwe (2020) predict below average matric outcomes to last 11 years without effective catch up.[53]

Maldonado and Witte (2020) estimate learning loss among primary school children in Belgium using six years of panel data fitted to a DD model. In particular, they identify the deviation in student outcomes on standardised tests in 2020 from the previous time trend, finding that inequality within and between schools rose following closures, and that schools with a more disadvantaged student population experienced greater learning losses. Unlike Ardington et al. (2021), however, Maldonado and Witte (2020) do not find evidence that those at the bottom of the standardised test score distribution experience greater losses in learning.

Angrist et al. (2020) show learning gains from a low-tech intervention that promoted continued learning during school closures in Botswana. Through a rapid randomised control trial,

[52] For more information on the responses of all universities, van Schalkwyk (2021) reflects on the sector-level response to COVID-19, Chaka (2020) provides a review of the teaching and learning responses to COVID-19 closures of 21 of the 26 public universities in South Africa, and Universities South Africa (2020) provides a summary of institutions' responses in the early weeks of the lockdown.

[53] See Azevedo et al. (2020) for global simulations of learning loss due to school closures, as well as the estimated cost in terms of earnings.

the authors study the effects of text messages and phone calls as a substitute to schooling during the COVID-19 pandemic. Moreover, the authors find high demand for low-tech interventions continuing after the trial period. However, Angrist et al. (2021) suggest half to over a year's worth of learning loss using Early Grade Reading Assessments in Ethiopia, Kenya, Liberia, Tanzania, and Uganda.

One of the key features of these school-level studies is the use of standardised assessments that allow the authors to identify learning loss. Since there are no standardised exams administered at undergraduate level at UCT,[54] however, I note that the DD approach I employ in this chapter cannot draw an explicit association between changes in academic performance with learning gains or losses.

Early research on the effect of the pandemic among university students along income gradients can be found in Aucejo et al. (2020). The authors survey students from one of the largest public institutions in the United States [US], with the survey explicitly designed to recover counterfactual outcomes in the absence of the pandemic. For example, the authors ask students about their current GPA in a post-COVID-19 world, as well as their expected GPA in the absence of COVID-19. They thus obtain a subjective treatment effect of COVID-19 on academic performance. In particular, the authors find that lower-income students were 55% more likely to delay graduation due to COVID-19 than their higher-income peers.

Rodriguez-Planas (2020) similarly finds stronger effects of the pandemic among low-income students. Using administrative data merged with a student survey, they document the educational, financial, and personal burdens of low-income public university[55] students from an urban college with an ethnically diverse student population in the US. They find that low-income college students were 8% more likely than the general student population at the same college to experience challenges with online classes. This owed mostly to greater childcare responsibilities, lack of internet, or increased probabilities of experiencing distress.

Rodriguez-Planas (2021) further considers the impact of the COVID-19 pandemic on university student academic performance, as measured by credits dropped and earned, as well as grade point averages [GPA]. Drawing on panel data of students from the same college mentioned above, the author estimates sharp increases in the Spring 2020 GPA for the sample relative to earlier semesters. In addition to the potential contributors to increased performance that I identify in the introduction to this chapter, Rodriguez-Planas (2021) notes that a flexible grading policy at the institution in their study may have contributed to the changes in outcomes observed.

By classifying students into low- and high-income categories (based off ever receiving the

[54] There is no country-wide standardised assessment for university students in South Africa. An exception is the final exam of the Post-Graduate Diploma in Accounting. However, I do not believe that the subset of students who write this exam is generalisable to the university student body as a whole.

[55] Queen's College – a four-year college in the New York City's public university, City University of New York.

federal Pell grant[56]), as well as quartiles of previous academic performance (based on GPA in the previous semester), Rodriguez-Planas (2021) finds that students in the bottom quartile of the performance distribution dropped fewer credits and earned greater credits during the Spring 2020 semester relative to earlier semesters. That the increase in performance is inversely related to pre-pandemic performance is not surprising according to the author, since there is more room for improvement at the bottom of the distribution. Incorporating income status effects, the author finds top-performing lower-income students experienced a decrease in both grades and earned credits during the Spring 2020 semester relative to their higher-income peers. In contrast, lower-income students in the bottom quartile of the performance distribution outperformed their higher-income peers in Spring 2020.

Rodriguez-Planas (2021) notes that differential treatment (e.g. in marking leniency) by students' income status is highly unlikely, since income status is not observed. Thus, although changes in marking and/or cheating may be behind better performance in the whole sample, other factors may drive differences in outcomes by income status. For example, it is suggested that lower-income top performers may have struggled more with online learning than their higher-income peers. On the other hand, the author suggests that concerns about retaining financial aid may have driven this performance differential between high-income and low-income students at the bottom of the performance distribution (good academic performance is often a requirement of continued aid).

Lastly, research finding improved performance during university closures in Spain is documented in Iglesias-Pradas et al. (2021) as well as Gonzalez et al. (2020). Iglesias-Pradas et al. (2021) acknowledge that cheating may have driven their results, and Gonzalez et al. (2020) attribute improved academic outcomes to more efficient learning strategies during the lockdown, since the format of the assessments considered in their study did not change.

The existing literature on the effects of closures on tertiary education aligns broadly with findings of improved student performance in 2020, documented in UCT's (2020) Teaching and Learning Report: one of the best sources currently available on what happened at UCT during COVID-19. I expand on the contents of this report in the following section on institutional context.

3.3. Institutional context and COVID-19-related closures

The University of Cape Town is one of South Africa's 26 public universities. It is a historically-advantaged, traditional university,[57] with the consequence that the institution itself,

[56] A form of financial aid in the US that is awarded to undergraduate students who display exceptional financial need and have not earned a bachelor's, graduate, or professional degree. See more about the Federal Student Aid System here

[57] In South Africa, three types of universities exist: traditional universities (research-intensive), comprehensive universities (mixed research and teaching focuses), and universities of technology (predominantly teaching institutions). Historical (dis)advantage reflects apartheid legislation on the

as well as its student body, are relatively more affluent than student bodies at other institutions. Nonetheless, the university enrols a wide demographic of students in line with its admissions policy to redress racial disadvantage. UCT has an explicit internal financial aid program for academically eligible students who are unable to pay tuition fees, thus further broadening its scope to enrol a diversity of students.[58]

Prior to 2018, UCT ensured that students with a household income up to R230 000 per annum were covered by a combination of NSFAS funding and institutional funding – awarded on the same conditions as NSFAS funding.[59] UCT would also top-up this funding to ensure students' full cost of study was covered. As such, when NSFAS funding is referred to in this chapter, it should be understood as the combination of NSFAS and UCT funding.

After the increase in the NSFAS financial eligibility threshold for new students in 2018, UCT undertook to fund all students whose household income was below R350 000 under the same conditions of award of the new policy. This included those enrolled prior to 2018. This differs from the national policy, which funds those enrolled prior to 2018 only if their household income is below R122 000. UCT additionally ensures that students whose household income is too high to qualify for NSFAS funding, but too low to afford fees, are offered a combination of UCT loans and bursaries. This funding is termed 'GAP' funding. Students from families with combined income up to R600 000 per annum are eligible for GAP funding, with this income threshold having increased on a sliding scale over the years.

Before COVID-19-related closures, the predominant teaching modality at UCT was face-to-face (contact learning). However, the institution was relatively productive in its utilisation of online and 'blended' learning techniques during this time.[60] This set it up to move the academic year online relatively quickly and smoothly after the campus closed following the implementation of national lockdown on 26 March 2020. Online, 'Emergency Remote Teaching' [ERT] started approximately one month after the initial lockdown was announced, with a period of online orientation preceding that. ERT was to be asynchronous and designed for low bandwidth and limited internet access (UCT, 2020).

The university provided laptops on loan to all students who were eligible for or were funded

classification of institutions. UCT is considered historically advantaged given that it was historically designated to serve white South Africans during apartheid.

[58] According to data in UCT's (2020) Teaching and Learning Report, close to one quarter of the 2020 undergraduate entrance cohort came from families within which educational and social disadvantage was severe. Additionally, of the 2 205 UCT respondents to DHET's Students' Access to and Use of Learning Materials [SAULM] survey, 61% were first-generation students (Centre for Teaching and Learning (UFS), 2020).

[59] I.e. income-contingent loans.

[60] Student protest action across universities in 2015 and 2016 necessitated a mixture of contact and distance (online) 'blended learning' (Czerniewicz, 2020), which likely contributed to UCT's ongoing utilisation of technology to supplement in-person lectures.

through financial aid programs. This meant these students were not subject to the lengthy delays on the delivery of laptops that NSFAS-funded students at other institutions faced (Maqhina, 2020). Like other universities in the country, it was negotiated for the institution's online sites and e-learning platform to be zero-rated, and students additionally received data bundles that they could use to access online resources. For students who could not access the internet in any form, UCT distributed printed learning materials and USB drives so that students could keep up with their courses (UCT, 2020). During 2020, as lockdown regulations were revised, UCT allowed for the phased return of the most vulnerable students to campus (UCT, 2020). Online learning, however, continued for the remainder of 2020.[61]

According to results presented in the UCT Student Experience Survey report (Marquard et al., 2020), one-third of student respondents indicated that their living circumstances were not suitable for studying. Additionally, 34% felt that they did not have adequate time for their studies, and 33% reported financial stressors (UCT, 2020). Findings further indicated that students felt that there was too much course content for the hours allocated. As a report on Students' Access to and Use of Learning Materials [SAULM] Survey indicates, some students felt that courses provided them with a lot of material, but no clear direction on how to engage with the material (Centre for Teaching and Learning (UFS), 2020).

Nonetheless, the Data Analytics for Student Success Project at UCT reported improvements in students' marks. There are a number of factors thought to have contributed to this, including accounts of marker leniency and student collusion, but also improved quality of engagement from both students and staff, better responsiveness to students' needs and perhaps an improved quality of assessment, given the move from more 'traditional' exams (UCT, 2020). Regarding the latter, anecdotal accounts suggest that a move towards continued assessment removed students' stress around a single exam affecting course performance, but the pressures of their workload increased. UCT's (2020) Teaching and Learning Report also suggests that asynchronous teaching may have supported students by allowing a different pace of learning in which proficiency in English was less of an issue than in face-to-face classes.

Furthermore, the cancelling of Duly Performed [DP] requirements in 2020 may have contributed to improved student performance, especially for struggling students who may ordinarily be filtered out by DP requirements before they write exams or the final assessment (in essence accumulating 0 additional credits for each course in which DP is enforced). In 2020, however, all students had an equal chance to succeed by the end of the course.[62] The easing of pre-requisites and suspension of academic exclusions for 2020 reflected decisions taken in support

[61] More information about the institution's response to COVID-19, together with details of student support offered and other institutional surveys, can be found in the 2020 Teaching and Learning report here.

[62] If this is a factor contributing to improved performance, it could imply that some students who usually stumble at the DP hurdle have the required knowledge to pass.

of UCT's commitment not to leave any student behind (Lange, 2021).

The institution concluded the 2020 academic year in December 2020, and in 2021 it adopted a physically distanced, low-density campus approach, with certain programs allowing students to resume contact learning and examination. All programs were to plan for their students to receive some form of contact instruction on campus on a rotation basis, and residences were to operate at full capacity. The institution also made available study spaces for students who did not live in residences, but who were negatively impacted by structural and systemic inequality. During 2021, students' work loads were also expected to return to pre-COVID-19 levels – 45 hours per week vs. the 30 hours per week expected in the previous year (UCT, 2020). A 2021 report to Senate (Lange, 2021) notes that since the beginning of the 2021 academic year, students were observed to be struggling with their work loads, academics observed a lack of engagement, and a preliminary analysis of marks showed uneven performance.

3.4. Conceptual foundations

In this section, I introduce a theoretical framework with the intention to unpack pathways through which student performance may have shifted, in order to guide the interpretation of results in the following section. To begin, I present a utility model of student performance in a pre-pandemic reality, thereafter describing anticipated shifts in the parameters of the model due to COVID-19-related closures.

I draw on the work of Devadoss and Foltz (1996),[63] who propose a straightforward model following Becker (1965), premised on the assumption that a utility maximising student is both a consumer and a producer of educational outcomes.[64] The utility framework presented by Devadoss and Foltz (1996) is well-suited to this context, since the authors study the effects of student behaviour (e.g. time spent on education), teacher attributes (e.g. marking leniency) and course characteristics (e.g. contact or remote learning) on educational performance. The model is presented below.

3.4.1. A utility maximising student

Assume the utility function of a student is given by:

$$U = U(C, R, E), \tag{3.1}$$

where C is a composite commodity, R is leisure, and E is educational performance. Educational performance depends on market and non-market factors given by the educational production function:

$$E = E(S, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{F}), \tag{3.2}$$

[63] And to a lesser extent that of Levin and Tsang (1987).

[64] The model extends neo-classical utility theory by incorporating an educational production function. Becker (1965) proposed this concept more generally when he suggested that households are both producers (through combining inputs of goods and time) and utility maximisers.

where S is time spent on education, $\boldsymbol{X}$ is a vector of inputs[65] such a textbooks and laptops, $\boldsymbol{Z}$ is a vector of endowment factors such as motivation and innate ability, and $\boldsymbol{F}$ is a vector of other learning services and educational factors including the nature of the course and assessments, as well as instructors' teaching skills and attributes.

Endowment factors, $\boldsymbol{Z}$, can be subsumed in the educational production function (see Becker, 1965), meaning equation 3.2 can be equivalently presented as:

$$\begin{aligned} S &\equiv sE, \\ X_i &\equiv d_i E, \\ F_i &\equiv a_i E, \end{aligned} \tag{3.3}$$

where s gives the input of time per unit of E, and d_i and a_i give the amount of input X_i (a component of $\boldsymbol{X}$), and other factors F_i (a component of $\boldsymbol{F}$) required per unit of educational output E, respectively. A student then faces the following budget and time constraints:

$$I + WL = P_C C + \Sigma P_{X_i} X_i + \Sigma P_{F_i} F_i, \tag{3.4}$$

$$T = R + S + L, \tag{3.5}$$

where I is endowment income (e.g. financial support), W is the wage rate, L is labour hours, P_C is the price of the composite commodity, P_{X_i} is the price of input X_i, and P_{F_i} is the price of educational or learning service F_i. Income is allocated between consumption, C, and other educational inputs and services, X_i and F_i. Time, T, is allocated between leisure, R, time spent on education, S, and work hours, L.

Following Becker (1965), equations 3.3 - 3.5 can be combined into a single resource constraint:

$$I + WT = P_C C + (\Sigma P_{X_i} d_i + \Sigma P_{F_i} a_i + Ws)E + WR.$$

Full income, $I+WT$, is spent directly on the composite commodity, C, and indirectly through foregone income by spending time on education and leisure.[66] Students maximise their utility subject to this resource constraint. In principle, for a specific utility function, the first-order conditions of the following utility maximisation problem can be solved for C, R and E:

$$U(C, R, E) + \lambda[I + WT - P_C C - (\Sigma P_{X_i} d_i + \Sigma P_{F_i} a_i + Ws)E - WR] = 0.$$

3.4.2. Campus closures: shifting parameters of utility maximisation

Before the pandemic hit, on-campus learning and living typically presented a way – albeit imperfect – of equalising access to resources for students from differing backgrounds. As

[65] These can also be considered a level of learning resources (see Levin & Tsang, 1987).

[66] Becker (1965) describes full income as "the sum of money income and that forgone or "lost" by the use of time and goods to obtain utility".

Rodriguez-Planas (2020) notes, the pandemic suddenly changed the environment (both institutional and economic) on which many students depended to perform academically. That is, campus closures increased the relevance of the home environment to the educational environment. Thus, the closure of campus and student residences will have shifted a number of parameters of a student's utility maximisation problem described above.

First, it is evident that the educational production function (equation 3.2) will have shifted due to the changing nature of courses and assessments (i.e. the move to online and distance learning from face-to-face contact). For example, if online platforms made it more interesting for students to learn, meant a more productive use of technology, or facilitated equity in learning, educational performance may have improved. On the other hand, lack of access to resources for remote learning in home environments will likely have adversely impacted students' educational production functions – that is, if educational inputs (the X_i) required for remote learning are unavailable to students, students' educational performance may suffer.

Second, the educational production function will shift with changes in marker leniency. Anecdotally, it has been suggested that marking leniency increased in 2020. That is, markers may have faced altered incentives to grade more leniently or to inflate grades. An awareness about widespread student-body inequalities and the household environments to which many students had to return may have encouraged a sympathetic attitude toward students at the time, prompting a lenient approach to grading. Additionally, where time was a scarce commodity during COVID-19 lockdowns, constraints on staff and the opportunity costs of their time (e.g. research activities, home care responsibilities) may have preempted lenient grading. Lastly, difficulties in delivering remote teaching of the same quality achieved in person may have encouraged leniency toward students among teaching staff.

In addition to shifts in the educational production function, students' time constraint (equation 3.5) is likely to have shifted. Since students were required to return (and/or remain) home, instead of devoting time to education, they may have been required to provide help and/or care at home, or find work to support their household in the face of economic hardship. Moreover, students' budget constraints would have shifted if income (or its sources) changed. Students may have been financially supported by family who could no longer afford to do so, continuation of financial aid may have been a concern, or they may have lost labour market income from jobs near to campus that were forfeited upon returning home. Shifts in endowment income may have lead to some students picking up part-time work to self-fund their studies.

Due to difficult and uncertain learning conditions, students may have faced incentives to cheat. There are a number of reasons why the move to online learning may have facilitated an increase in academic dishonesty, over and above the fact that students may have been struggling academically,[67] and the fact that in a remote assessment context students may believe they are

[67] If a student is already failing a course, then penalties for cheating may not be such a deterrent. Indeed, Carrell et al. (2008) suggest peer influence on cheating is greater among lower performing students.

less likely to get caught cheating.[68]

Augusta and Henderson (2021) note that if a course is viewed as a product for which the student is paying, then students may have felt they were not receiving what they paid for when courses switched from in-person to online. Consider that in the budget constraint (equation 3.4), the price of learning/educational services (e.g. university fees) will not have changed, but the purchased 'products' or services, the F_i, will have. If students themselves feel cheated by this, they may feel justified in their cheating. Relatedly, Augusta and Henderson (2021) believe that if the perception that the quality of online education is low, students may be less likely to invest in learning and hence be more likely to participate in dishonest behaviour and cheating.

While evidence of increased academic misconduct among university students in South Africa during remote assessments is largely anecdotal, it has been noted that the number of academic misconduct cases being referred to the University Student Disciplinary Tribunal at UCT increased in 2020 from previous years (Mafolo & Shoba, 2021).[69] A survey of academic staff employed across South Africa's universities also suggests that 73% of respondents raised concerns about academic misconduct (Nkosi, 2021).[70]

My *a priori* expectation on the overall effect of closures and the implementation of online learning and assessment is uncertain owing to the wide array of interacting and inseparable factors described above which will have influenced students' academic performance during this time. Moreover, there are a number of factors that I mention above that are not observed in my data – marker leniency and cheating being obvious examples. This means that although it is useful to conceptually consider various pathways through which academic performance in 2020 would have been affected, these mechanisms are not testable with data that I have access to. My empirical strategy, while guided by the above framework, thus focuses on describing the changes in student performance that occurred and how they may have impacted existing inequalities within the system, rather than attempting to uncover which behavioural responses underpinned such changes. Nonetheless, this conceptual framework provides suggestive evidence of which factors are likely to have contributed to changes in academic performance during the pandemic, and motivates that the extent to which shifts in incentives and the parameters of utility maximisation will be conditional on factors such as the home environment and funding considerations.

[68] Especially if sharing online is harder to observe than physical copying. Leung (1995 in Hutton, 2006) suggests that a greater *likelihood* of punishment has a greater deterrent effect on cheating than does a more *severe* punishment.

[69] This is not to say all cases were a result of intentional dishonesty. Some students may have lacked knowledge/understanding about how to use online resources in an ethical way.

[70] This is not unique to South Africa. Augusta and Henderson (2021) present data from a large Canadian university on trends in student academic misconduct cases. They particularly highlight a spike in cases generated from online subscription based tutoring services used during final exams in 2020.

3.5. Data, performance measures, and empirical strategy

3.5.1. Data

This study draws on annual, individual-level institutional data on students enrolled in their first undergraduate qualification at UCT between 2016[71] and 2021 in all programs of study.[72] I use data from UCT, rather than the national-level data that is captured for all universities, since the UCT data is more comprehensive in the information that it captures about students compared to the corresponding data in the national database. For example, the UCT data contains measures of Grade Point Average [GPA], information on students' matric (school-leaving) results, as well as information on which students receive financial aid. This information is lacking in the national data but is central to the empirical strategy. Although the experience of students attending UCT will not reflect the national picture, this study provides insight into the experience of students at one of the better-resourced institutions in South Africa. Table 3.1 shows the mean characteristics of undergraduate students entering university for the first time between 2016 and 2021.

Just over half of each entry cohort identify as female, with the majority of students enrolling in the faculties of Commerce, Humanities, or Science, respectively. Although it appears that the share of White, Coloured, and Indian students has been declining over time, there is an increasing trend in students not specifying their race. The share of students on NSFAS financial aid has been increasing over time, where the observed decrease in the share of students on GAP funding from 2018 onward corresponds to the increase in the financial eligibility threshold for NSFAS funding. Thus, some students who were previously on GAP funding would fall under NSFAS funding instead.

Matric aggregate is the summation of students' six matric-subject marks (percentages), excluding life orientation. Where students have completed more than six subjects, the score is calculated using the result for English and any required subject(s) for the relevant program, along with the remaining highest scoring subjects. Just over half the students who are missing a matric aggregate list their home country outside of South Africa (not shown here), indicating that a portion of missing information in this case may owe to students not completing their secondary education in South Africa.

3.5.2. Credit loads and credits passed as academic performance measures

The academic performance measures of interest are credits taken, credits passed, and the share of credits passed to taken. In South Africa, university qualifications (and their respective courses)

[71] The 2016 cohort is chosen as the first entry cohort in the analysis since prior cohorts were enrolled under a different admission's policy.

[72] I am grateful to Jane Hendry, Chief Information Advisor – Institutional Information Unit, Institutional Planning at UCT, for facilitating access to this data and for generously offering her time to share insights along the way. I am also grateful to Tasneem Salasa, Director of Student Financial Aid at UCT, for her assistance with the financial aid data.

Table 3.1 Average student characteristics by cohort

Cohort	2016	2017	2018	2019	2020	2021
Year of study[a] in 2020	5	4	3	2	1	-
Female	0.54	0.54	0.55	0.55	0.56	0.54
Race						
White	0.26	0.20	0.17	0.16	0.16	0.11
Black	0.33	0.35	0.34	0.33	0.35	0.31
Chinese	0.01	0.01	0.00	0.00	0.00	0.00
Coloured	0.17	0.16	0.14	0.11	0.12	0.10
Indian	0.08	0.07	0.05	0.05	0.05	0.04
Not specified	0.15	0.21	0.29	0.35	0.30	0.44
Faculty						
Commerce	0.23	0.21	0.23	0.22	0.20	0.23
EBE	0.14	0.12	0.12	0.10	0.08	0.07
Humanities	0.27	0.30	0.31	0.33	0.35	0.30
Law	0.04	0.04	0.03	0.04	0.03	0.01
Health Sciences	0.12	0.11	0.08	0.07	0.07	0.07
Science	0.20	0.21	0.23	0.24	0.27	0.32
Financial aid						
NSFAS+UCT	0.20	0.27	0.33	0.36	0.34	0.37
GAP	0.06	0.06	0.03	0.02	0.02	0.01
Grade point average[b]						
Year 1	61.13	60.41	59.61	60.05	-	59.84
Year 2	59.59	58.69	59.88	-	57.61	-
Year 3	59.78	60.64	-	57.70	-	-
Year 4	60.67	-	58.45	-	-	-
Year 5	-	-	-	-	-	-
Matric aggregate	466.05	452.29	449.31	459.75	457.85	467.26
Share missing matric aggregate	0.12	0.12	0.09	0.10	0.09	0.10
Student-year observations	16 115	15 021	12 632	11 131	7 715	4 141
No. students in entry cohort	4 176	4 023	3 596	3 896	3 947	4 141

Source: Author's own calculations using UCT (2016-2021).
Notes: [a] Year of study refers to year since start. [b] GPA was not calculated in 2020.

comprise a certain number of credits, where credits are intended to reflect a measure of the volume of learning required for a qualification or course. Credits are quantified as the number of notional study hours required for achieving the learning outcomes for qualification or course, and ten study hours are rated as equivalent to one credit.[73] Since Certificate, Diploma, Bachelor's Degree, and Bachelor (Honours) Degree qualifications assume a 30-week full-time academic year, an average full-time equivalent student is expected to study for a 40-hour week, thus requiring a minimum credit load of 120 credits per year. It follows that the minimum total credits for a three-year Bachelor's degree be 360 credits, and a four-year Bachelor's 480 credits. Credits can thus represent an important measure of students' academic progress while they are enrolled. Despite shifts in the number of study hours expected in 2020 (a decrease from 40 per week to 30 per week), the credit weighting of students' courses and qualifications did not change.

[73]Note that the same course may be weighted by a different credit value depending on the qualification for which the student taking the course is enrolled.

The number of credits comprising a qualification is governed by the Higher Education Qualification Sub-Framework [HEQSF], which "establishes common parameters and criteria for qualifications design and facilitates the comparability of qualifications across the system" (Council on Higher Education (South Africa), 2013). Nonetheless, the HEQSF documentation notes that within the common parameters, program diversity and innovation are nonetheless encouraged. That is, institutions have a broad scope within which to design educational offerings that are aligned to their different visions and missions, and to meet the varying needs of their stakeholders and communities. As such, the minimum qualification credits at UCT may differ from that offered at other institutions, despite falling under the same HEQSF. At UCT, for example, a four-year Bachelors comprises 601 credits on average, and a three-year Bachelors comprises 418 credits on average.[74] Table B.1 in Appendix B summarises undergraduate qualification credits at UCT by minimum time for degree completion and faculty.

The choice of credit-based measures is motivated by the reasoning that they should, in theory, reflect a university-wide measure of learning. Together with student dropout (discussed later), looking at credits taken and passed allows me to assess whether students adjusted their enrolment and learning (credit) load in response to the conditions in 2020. These measures, however, simply indicate whether or not credits were taken and passed, and measures of differences in performance conditional on passing are not available.

Table 3.2 shows the average values of credit measures by cohort and year of study. The year of study in which a cohort is affected by the pandemic is highlighted in grey. The number of credits taken fell marginally in all years of study in 2020, except for the 2020 entering cohort. Credits passed, on the other hand, increased at all levels of study. In 2021, credits taken increased to be on par with pre-pandemic levels, but credits passed fell below pre-2020 levels.

Among second-year students, the share of credits passed has been improving over time, albeit marginally. It is thus not immediately straightforward to conclude that performance improvements in 2020 would not have occurred in the absence of the pandemic. To plausibly measure the impact of the pandemic on these measures, one therefore needs a counterfactual against which to compare performance. I employ difference-in-differences[75] [DD],

[74] This data on the total credits required for UCT qualifications was matched to the UCT data from the Higher Education Management Information System [HEMIS] data, captured at the qualification level (See Figure B.1 in Appendix B). However, qualifications have different qualification specialisations (e.g Bachelor of Science *in...*) which, according to UCT undergraduate handbooks, can differ from each other in their total required credits. E.g. a Bachelor of Business Science in Marketing requires 695 credits, but a Bachelor of Business Science in Economics requires a minimum of 631 (Faculty of Commerce, 2022). In the HEMIS data, a Bachelor of Business Science is captured as requiring 612 credits. This specialisation-specific credit information is not available in the data I have access to. As such, I do not consider students' credit progress toward their total qualification credits.

[75] In fact, given that I draw on more than two time periods, the terminology 'Two-Way Fixed Effects' [TWFE] is more appropriate.

Table 3.2 Average performance measures by cohort and year of study

Cohort	2016	2017	2018	2019	2020	2021
Year of study[a] in 2020	5	4	3	2	1	-
Credits taken						
Year 1	142.21	141.24	146.03	146.81	147.07	148.51
Year 2	149.87	149.33	150.52	148.74	149.39	
Year 3	155.93	153.49	152.65	154.65	-	-
Year 4	141.07	137.80	141.24	-	-	-
Year 5	131.48	131.68	-	-	-	-
Credits passed						
Year 1	123.16	122.14	124.93	126.35	132.33	123.97
Year 2	127.94	127.94	130.68	133.96	121.34	-
Year 3	136.89	135.49	140.78	126.61	-	-
Year 4	125.20	127.35	116.33	-	-	-
Year 5	118.90	104.18	-	-	-	-
Credits passed to credits taken (ratio)						
Year 1	0.86	0.86	0.85	0.85	0.89	0.83
Year 2	0.84	0.85	0.86	0.89	0.80	-
Year 3	0.87	0.87	0.92	0.81	-	-
Year 4	0.88	0.92	0.82	-	-	-
Year 5	0.89	0.78	-	-	-	-
Student-year observations	16 115	15 021	12 632	11 131	7 715	4 141
No. students in entry cohort	4 176	4 023	3 596	3 896	3 947	4 141

Source: Author's own calculations using UCT (2016-2021).
Notes: All outcome measures are conditional on enrolment in a given year. [a] Year of study refers to year since start.

a quasi-experimental approach, and compare changes in performance for students during the pandemic against performance of their peers prior to the pandemic. I describe this approach below.

3.5.3. Empirical strategy

Academic performance in 2020

To plausibly measure the impact of the pandemic on these measures, one needs a counterfactual against which to compare performance. I employ difference-in-differences[76] [DD], a quasi-experimental approach, and compare changes in performance for students during the pandemic against performance of their peers prior to the pandemic. Following Ardington et al. (2021), I estimate the DD using an Ordinary Least Squares [OLS] regression of the form:

$$Y_{it} = \beta_0 + \beta_1 C_i + \beta_2 T_{2020} + \beta_3 (C_i \cdot T_{2020}) + \gamma_i + \epsilon_{it}, \tag{3.6}$$

where Y_{it} is the academic performance measure for student i in calendar year t, C_i is a indicator variable for whether the student's cohort is in year 2-4 in 2020[77], T_{2020} is an indicator

[76] In fact, given that we draw on more than two time periods, the terminology 'Two-Way Fixed Effects' [TWFE] is more appropriate.

[77] I.e. 2016 is not a COVID-cohort since they would be in year 5 in 2020. Similarly, the 2020 entrance cohort is not included in this analysis.

for 2020, and ϵ_{it} is an idiosyncratic error. The individual fixed effect, γ_i, is removed through estimating within-individual effects (i.e. a fixed-effects estimator).[78] The DD estimate is thus obtained as the β_3 coefficient, interpreted as the differential impact of the 2020 year on students' performance.

I take note here of recent developments in the DD and event study literature that account for heterogenous treatment effects and variation in treatment timing (Callaway & Sant'Anna, 2021; Goodman-Bacon, 2021; Sun & Abraham, 2021). A concern highlighted in this literature is that if there is variation in treatment timing, a simple OLS regression may make 'forbidden comparisons' (Duhaut et al., 2021a, 2021b), in which groups that are treated earlier in time act as controls for groups that are treated later. However, since COVID-19 affected all cohorts in the study at the same point in time, and for the same duration of time, no forbidden comparisons are made in this analysis.

The DD strategy relies on the assumption that, in the absence of the pandemic, the cohort affected by COVID-19 in a given year of study would have experienced a counterfactual outcome identical to the observed outcome in cohorts that were not affected by COVID-19 in the same year of study. For example, in 2019, the 2018 cohort would be in their second year of study. Whereas for the 2019 cohort, second year takes place in 2020. In the absence of the pandemic, the assumption reflects the expectation that the second-year performance for the 2019 cohort would have followed the same trend as those of the 2018 cohort. While DD allows for the cohorts to be observationally different, it assumes that this difference is constant over time and can be differenced out.

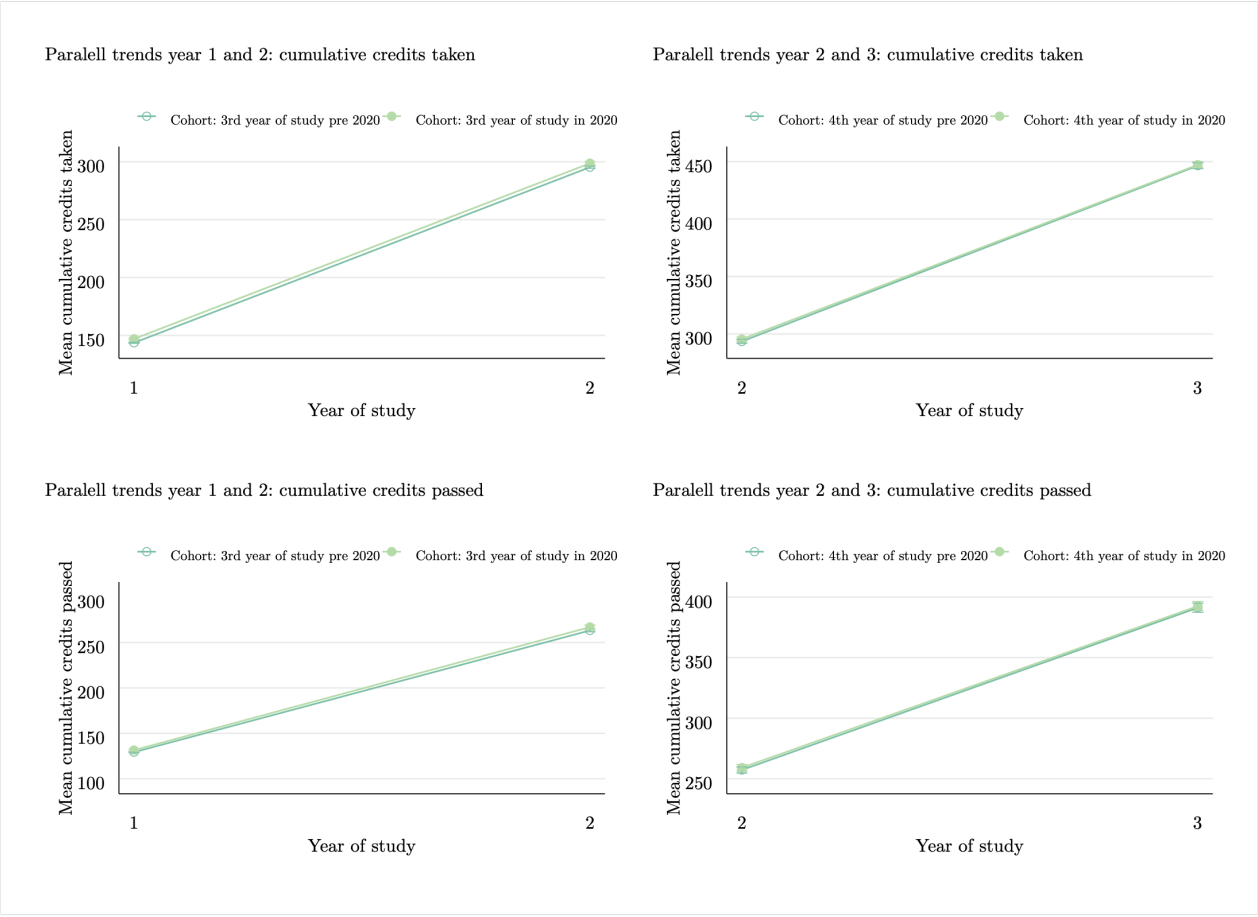
Figure 3.1 provides a visual inspection of parallel trends, year by year. For both credits passed and credits taken, trends are not significantly different between years 1 and 2 for the cohort whose third year of study falls in 2020, and those whose third year occurred before 2020. This similarly holds true for the cohort in fourth year in 2020. I do not examine pre-trends in first year performance, give the lack of a previous period.

The year-on-year nature of the trends in Figure 3.1 draws attention to the fact that the specifications use all pre-COVID years of study as controls for COVID-19 years of study, irrespective of the year of study. That is, pre-COVID trends in year 4 credits taken/passed, as well as trends in years 1, 2, and 3 credits taken/passed will act as controls for year 4 in 2020. As such, results should be interpreted as the average effect of COVID-19 across all years of study, with years of study with more observations weighted more heavily in the control group (e.g. year 2). Since trends in credits taken and passed may be different for different years of study, I estimate equations 3.6 and 3.7 including controls for year of study.[79]

[78] This allows me to control for the fact that credits passed is a function of the number of credits taken, which will be influenced by students' individual history.

[79] I also test the sensitivity of these results to how year of study is defined. In the main analysis, it is defined by number of years since the student starts, but I confirm that results are substantively similar

Figure 3.1 Pre-trends by outcome measure and year of study



Source: Author's own calculations using UCT (2016-2020).
Notes: Sample comprises those in entrance cohorts 2016-2019.

Differential effects by GPA and financial aid status

Given the findings of existing literature discussed in Section 3.2, I expect that students from lower socioeconomic backgrounds may have been affected more severely by the pandemic. I also expect that students' performance in 2020 may have been related to prior academic performance. I thus include interactions between C, T, GPA quartile of the previous year's cumulative GPA distribution (hereafter reference to quartiles refers to quartile of the pre-pandemic GPA distribution), and an indicator variable for whether or not the student is a recipient of NSFAS funding – a proxy for SES.[80] Note, however, that individual fixed-effects are not differenced out when including GPA quartile and NSFAS interactions. I test the robustness of these results to

when defining year of study by year of enrolment. The difference in definition occurs where year since start will classify a student who is not continuously enrolled as year 3 if they took a year off after first year, whereas year of enrolment will classify them as being in year 2.

[80] Since being on financial aid in 2020 and student performance may be correlated, I also consider differences along quartiles of the matric performance distribution. The matric examination process is exogenous to the university system and thus less likely to be related to continued financial aid.

the inclusion of variables that capture students' backgrounds.

Another relevant consideration is students' faculty. UCT's (2020) Teaching and Learning Report notes that faculties such as Humanities, that had already replaced exams with continuous assessment, experienced fewer assessment problems during 2020 than Science, Commerce, Engineering and Built Environment [EBE], and Law, which typically relied on invigilated exams to ensure the integrity of assessments. Moreover, faculties differ in the programs they offer. That is, some faculties offer programs with a greater proportion of work that cannot be taught or properly assessed online. I thus explore faculty-specific differences in performance by running separate regressions of equation 3.6 for the six faculties at UCT.[81]

Academic performance in 2021

Recognising that academic experience during the pandemic could affect performance in subsequent years, next I examine academic performance in 2021. In the first instance, I am interested in 2021 performance relative to pre-pandemic trends. Performance may have improved, it may have returned to pre-pandemic levels, or it may have worsened. We estimate a version of equation 3.6, excluding the 2020 year:

$$Y_{it} = \rho_0 + \rho_1 P_i + \rho_2 T_{2021} + \rho_3 (P_i \cdot T_{2021}) + \gamma_i + \epsilon_{it}, \tag{3.7}$$

where Y_{it} is the academic outcome measure for student i in year t, P_i is a indicator variable for whether the student's cohort is in year 3-4 2021 (which we term the post cohort),[82] T_{2021} is an indicator for 2021 and ϵ_{it} is an idiosyncratic error. We remove the individual fixed effect, γ_i by estimating a within-individual fixed effects regression. The DD estimate is obtained as the ρ_3 coefficient, representing any differential shifts in performance in the year after the pandemic. Again, we further assess whether there are differences along the GPA distribution, [83] by financial aid status, and within faculty.

Second, I am interested in the extent to which potentially inflated performance in 2020 may be related to performance in 2021. I therefore include an interaction in equation 3.7 between $T21$ and an indicator variable for whether or not performance improved in the year prior.

Differential dropout

Lastly, student dropout and academic performance are likely to be correlated. Where dropout rates differ before and during the pandemic, this can introduce bias in the DD estimates. For the purposes of this analysis, dropout is defined by students who exit without completing their qualification, or who dropout during the year of interest. Students may, however, return in

[81] The six faculties at UCT are Health Sciences, Humanities, Law, Science, EBE, and Commerce.

[82] Again, this means the 2020 and 2021 entrance cohorts are not included in this analysis due to lack of a pre-COVID period.

[83] GPA distribution of two years prior, since GPA was not calculated in 2020 and 2020 performance should be omitted by definition.

subsequent years. Dropout rates by cohort and NSFAS funding status are shown in Table 3.3, where the dropout rate in 2020 is highlighted in grey (i.e. those who do not re-enrol after 2019 or who exit during 2020). Differential dropout between NSFAS-funded students and those not funded by NSFAS will have particular relevance for the analysis of achievement gaps (discussed in the following section) between students of different socioeconomic groupings.

Table 3.3 Dropout rate (percent) by cohort and year of study

	2016 Cohort		2017 Cohort		2018 Cohort		2019 Cohort		2020 Cohort		2021	
	NSFAS	Other	NSFAS	Other	NSFAS	Other	NSFAS	Other	NSFAS	Other	NSFAS	Other
Year 1	0.76	2.25	1.22	2.52	0.67	3.24	0.15	2.42	0.55	1.00	0.91	3.50
Year 2	7.03	9.43	7.07	9.00	7.39	9.19	5.20	8.67	2.80	6.77	-	-
Year 3	7.37	5.97	5.49	5.69	6.52	4.69	2.57	2.68	-	-	-	-
Year 4	6.49	7.76	7.79	7.20	2.25	2.11	-	-	-	-	-	-
Year 5	9.50	10.48	4.49	5.41	-	-	-	-	-	-	-	-

Source: Author's own calculations using UCT (2016-2021).
Notes: [a] Year of study refers to year since start. Dropout in year 1 reflects only those who exit during their first year of study. Year 2 reflects those who do not re-enrol after year 1 and who dropout during their second year.

Before the pandemic, the rate of dropout among students not funded by NSFAS was typically higher in all cohorts and years of study, with the exception of the 2016 cohort in year 3. This is consistent with the idea that NSFAS funding may provide students with an important safety net. During 2020, however, a greater share of NSFAS students dropped out in year 3 and 4 relative to students not funded by NSFAS, with the share of NSFAS-funded students dropping out also increasing relative to prior cohorts. However, a lower share of NSFAS-funded students in second year dropped out compared to those not funded by NSFAS, and this share was 2 percentage points lower compared to second-year dropout among NSFAS-funded students from previous cohorts.

If dropout represents one of the responses to conditions faced in 2020, the trends in Table 3.3 seem to suggest that students in different years of study may have implemented different strategies for coping with the uncertainty associated with the pandemic. If NSFAS-funded students in later years of study are unsure about continued funding if they under perform (i.e. they are closer to the allowable maximum years for funding eligibility), a strategic response may be to dropout. Indeed, I observe a 3 percentage point increase in dropout rates among fourth-year students in the bottom quartile (quartile 1) in 2020 relative to pre-2020 trends (from 14% to 17%; not shown here).

On the other hand, for second-year NSFAS-funded students in quartile 1, dropout rates fell by 4 percentage points in 2020 relative to pre-pandemic trends (from 18% to 14%; not shown here), while dropout rates increased from 26% to 31% among quartile 1 students who were not funded by NSFAS. In this regard, if students know that they have more years of NSFAS-funding remaining, NSFAS may be providing a considerable support structure to keep students enrolled.

Of particular relevance for the 2021 analysis is that dropout rates in 2021 are substantially

lower compared to previous years. This is, in part, likely a function of the cancelling of DP requirements during 2020 (see Section 3.3). To account for differential dropout in the analysis, I repeat the analysis both excluding dropouts, as well as including those who leave the university as taking and passing zero credits in the year following their last enrolment (i.e. the year in which they are recorded as dropping out). Note, however, that since I group years of study together in the analysis, some of the nuances described above (i.e. different dropout strategies by year of study) will not be apparent. The effect of differential dropout on the analysis is elaborated on in the following section.

3.6. Results and discussion

3.6.1. Academic performance in 2020

Table 3.4 presents the DD regression estimates for credits passed and credits taken by cohorts in years 2 to 4 of study in 2020.[84] The coefficient on the COVID effect represents the differential change in the average credit measure in 2020. For each measure, I show estimates from an OLS regression, a fixed effects [FE] regression, and the FE regression excluding dropouts.

Table 3.4 DD regression estimates

	Credits taken			Credits passed			Credits passed to credits taken		
	OLS	FE	FE (excl. dropout)	OLS	FE	FE (excl. dropout)	OLS	FE	FE (excl. dropout)
COVID effect (C·T)	-0.823	-3.268***	-4.173***	4.805***	2.461***	1.974***	0.046***	0.042***	0.041***
	(0.628)	(0.569)	(0.428)	(0.692)	(0.556)	(0.531)	(0.004)	(0.003)	(0.003)
COVID cohort (C)	0.468	-	-	0.560	-	-	0.002	-	-
	(0.597)	-	-	(0.820)	-	-	(0.005)	-	-
2nd year	-3.735***	-3.209***	6.064***	-3.546***	-2.971***	1.379***	-0.060***	-0.058***	-0.031***
	(0.385)	(0.380)	(0.252)	(0.392)	(0.377)	(0.357)	(0.002)	(0.002)	(0.002)
3rd year	4.565***	-1.630***	10.257***	6.782***	0.037	5.897***	-0.019***	-0.068***	-0.030***
	(0.538)	(0.525)	(0.364)	(0.620)	(0.529)	(0.491)	(0.004)	(0.003)	(0.003)
4th year	-11.315***	-17.900***	-2.307***	-6.824***	-10.888***	-2.778***	-0.025***	-0.067***	-0.009**
	(0.860)	(0.881)	(0.708)	(0.955)	(0.862)	(0.820)	(0.005)	(0.005)	(0.004)
Constant	140.618***	143.438***	144.419***	121.087***	123.768***	127.450***	0.836***	0.854***	0.879***
	(0.494)	(0.230)	(0.154)	(0.685)	(0.237)	(0.214)	(0.004)	(0.001)	(0.001)
R^2	0.01	0.04	0.04	0.01	0.01	0.01	0.01	0.02	0.02
Observations	48 611	48 611	45 883	48 611	48 611	45 883	48 607	48 607	45 879

Source: Author's own calculations using UCT (2016-2020).
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses and are robust to the presence of repeated observations on the same individual. The sample comprises those in entrance cohorts 2016-2019. In the OLS regressions, the COVID cohort coefficient shows the differential for the 2016 cohort, the only cohort not affected by COVID-19 by their fourth year of study.

In general, results show a decline in credit loads in 2020, with gains made in credits passed. Although significant, the performance gain in credits passed of 2.5 credits (in the FE regression)

[84] Although the 2016 cohort would be in their fifth year of study in 2020, year 5 is omitted since there is no control/comparison group in the data I have available.

is practically very small, given that a typical course is weighted 18 credits. Nonetheless, together with the decline in credits taken, the share of credits passed increased by a statistically significant 4.2 percentage points (in the FE regression) relative to pre-COVID trends.

Given the pandemic circumstances, students were granted leeway of 10 fewer study hours per week (a reduction from a 40 hour week to 30 hours in 2020), yet credit weightings for courses did not change. If students still had 40 hours available to dedicate to learning, the reduced hours expected per course may have enabled students to pass more courses, and therefore, credits. However, where credit weightings are intended to signal the amount of notional study hours, there is clearly a misalignment in 2020 between course learning outcomes, and credits awarded.

Overall dropout for students in years 2-4 in 2020 was slightly lower, on average, compared to pre-COVID rates. When students who dropout are excluded from the regressions,[85] I observe more muted effects of the pandemic on credits passed. This suggests that those who remained enrolled in 2020 performed slightly better than those who remained enrolled prior to 2020, but to a lesser extent than when differential rates of dropout are accounted for. The minimal decline in the *share* of credits passed, however, indicates that this is mainly a function of the reduction in credits taken. That is, those who remained enrolled in 2020 were taking even fewer credits than those who remained enrolled in pre-pandemic years. This suggests that some students responded to the pandemic in 2020 by adjusting their credit load rather than simply dropping out.

Since FE regressions are not feasible when including interactions with GPA quartile and NSFAS status, I highlight the similarity of the OLS and FE COVID effect estimates in the share of credits passed measure. In Table 3.5, I therefore consider the marginal effects of COVID-19 on the share of credits passed measure by GPA quartile and NSFAS status. Since I cannot eliminate individual fixed effects in these regressions, I show that the estimated size of the effect of 2020 is substantively similar when I include controls for students' gender, matric aggregate, and whether or not their first language is English (the medium of instruction and assessment at UCT).

It is apparent that for both NSFAS-funded students and students not funded by NSFAS, performance gains (in terms of the share of credits passed) decreased with GPA quartile. In particular, students not funded by NSFAS in quartile 1 experienced an increase in the share of credits passed by 9.5 percentage points, with NSFAS-funded students in the same quartile experiencing an increase of 8.8 percentage points. NSFAS-funded students in the top quartile experienced a decline in their performance, although this difference is not statistically different from pre-pandemic trends. Students in the top quartile who are not funded by NSFAS, on the other hand, experienced a significant increase of 1.5 percentage points on average, compared to non-NSFAS students in the same quartile prior to the pandemic.

For both NSFAS and non-NSFAS students, dropout is concentrated at the bottom end of the GPA distribution. When those who drop out are excluded, performance gains in quartile 1 become even larger. This magnified performance gain suggests that those who remained enrolled

[85] They are only excluded in the year in which they are not observed.

Table 3.5 Marginal effects of COVID-19 on the ratio of credits passed to credits taken by GPA quartile and NSFAS status

	Ratio[a]	+ Controls	Excluding dropout
Students not funded by NSFAS			
Quartile 1	0.095***	0.094***	0.134***
	(0.013)	(0.014)	(0.011)
Quartile 2	0.059***	0.068***	0.057***
	(0.008)	(0.008)	(0.006)
Quartile 3	0.021***	0.029***	0.020***
	(0.006)	(0.006)	(0.004)
Quartile 4	0.015***	0.016***	0.002
	(0.004)	(0.005)	(0.002)
NSFAS-funded students			
Quartile 1	0.088***	0.089***	0.097***
	(0.014)	(0.014)	(0.013)
Quartile 2	0.044***	0.050***	0.048***
	(0.011)	(0.011)	(0.010)
Quartile 3	0.024**	0.024***	0.017**
	(0.009)	(0.009)	(0.008)
Quartile 4	-0.009	-0.009	-0.007
	(0.011)	(0.011)	(0.008)
Observations	32 581	28 974	30 190

Source: Author's own calculations using UCT (2016-2020).
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses and are robust to the presence of repeated observations on the same individual. The sample comprises those in entrance cohorts 2016-2019. The regression with controls includes covariates for English first language, matric aggregate, and gender. The sample size reduces due to missing information for matric aggregate. First-year information is excluded due to lack of prior year GPA quartile.
[a] Full regression results are presented in Table B.2 in Appendix B.

in 2020 passed a substantially higher share of credits despite the fact that a likely greater proportion of weaker students remained (dropout is marginally lower in all quartiles in 2020 compared to pre-COVID years). The patterns in the other three quartiles are consistent with the overall picture discussed and presented in Table 3.4.

There are a number of factors that are likely to have contributed to the overall improvements observed, especially at the bottom end of the performance distribution. These include more room for improvement, accounts of marker leniency, and student collusion, but also improved quality of engagement from both students and staff, better responsiveness to students' needs, and perhaps an improved quality of assessment, given the move from more 'traditional' exams to continuous assessment (UCT, 2020). While improvements at the bottom end of the GPA distribution may have been artificial in a sense due to these factors, students' NSFAS status is typically not observable, and thus, at least in terms of marking, students' NSFAS status should have no impact on how leniently their assessments are treated.[86] That is, while it may

[86]This is especially expected to be the case for large undergraduate courses where often the markers employed are not the lecturers who taught the content and who were interacting with students directly. Marking is also done anonymously. This will, however, likely differ across faculties and course size. Also, if SES is linked to English proficiency, and markers are more lenient when English is less proficient,

be expected that the shrinking achievement gaps between stronger and weaker students could have partly been a function of factors such as marker leniency, better responsiveness to students' needs, and perhaps an improved quality of assessment, any change in the performance differential between NSFAS and non-NSFAS students in 2020 is unlikely to owe entirely to these factors.

To explore this further, Table 3.6 presents both pre-COVID and COVID achievement gaps between NSFAS-funded students and students not funded by NSFAS. Specifically, the top half of Table 3.6 shows the pre-COVID achievement gaps in credits taken, passed, and the share of credits passed, with the bottom half highlighting the respective gaps in 2020. Therefore, the 5.88 estimate in the quartile 1 row in the bottom panel, shows that among students in the bottom quartile, NSFAS funded students took 5.88 more credits than their unfunded peers.

Table 3.6 Differential in credit measures between NSFAS students and students not funded by NSFAS (achievement gap)

	Credits taken		Credits passed		Credits passed to credits taken	
	(1) Taken[a]	(2) Excl. dropout	(3) Passed[a]	(4) Excl. dropout	(5) Ratio[a]	(6) Excl. dropout
Pre-COVID gap						
Quartile 1	6.411***	-2.108*	-1.659	-8.360***	-0.003	-0.050***
	(1.772)	(1.156)	(1.764)	(1.669)	(0.012)	(0.011)
Quartile 2	-0.965	-5.105***	-7.097***	-10.852***	-0.023***	-0.047***
	(1.279)	(1.005)	(1.614)	(1.421)	(0.009)	(0.007)
Quartile 3	-1.733	-4.755***	-5.638***	-8.066***	-0.012	-0.028***
	(1.197)	(1.000)	(1.491)	(1.299)	(0.008)	(0.006)
Quartile 4	-0.632	-3.315**	-2.282	-4.867***	0.005	-0.012**
	(1.486)	(1.301)	(1.678)	(1.477)	(0.007)	(0.005)
COVID gap						
Quartile 1	5.882**	-5.180***	-5.147**	-16.145***	-0.010	-0.086***
	(2.407)	(1.572)	(2.410)	(2.255)	(0.016)	(0.014)
Quartile 2	-0.971	-3.387**	-8.265***	-10.999***	-0.038***	-0.056***
	(1.716)	(1.416)	(1.989)	(1.789)	(0.011)	(0.009)
Quartile 3	-3.163**	-6.505***	-7.383***	-10.696***	-0.009	-0.031***
	(1.513)	(1.362)	(1.738)	(1.608)	(0.008)	(0.007)
Quartile 4	-6.683***	-6.711***	-9.693***	-9.970***	-0.020**	-0.021***
	(1.781)	(1.524)	(1.963)	(1.745)	(0.009)	(0.006)
Observations	32 585	30 194	32 585	30 194	32 581	30 190

Source: Author's own calculations using UCT (2016-2020).
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses and are robust to the presence of repeated observations on the same individual. The sample comprises those in entrance cohorts 2016-2019. The table shows the marginal effects of NSFAS status on credit measures by GPA quartile and COVID-19 cohort. A negative effect indicates NSFAS-funded students are taking/passing fewer credits than their peers not funded by NSFAS. Regressions control only for year of study. [a] Full regression results are presented in Table B.2 in Appendix B.

then the assumption that SES is not observable to markers will not hold.

Considering the first column, in 2020 the gap in credits taken between NSFAS-funded students and students not funded by NSFAS grew in the upper two quartiles. On inspection of the data, it appears that a combination of increased credit loads among students not funded by NSFAS and a decrease in credit loads among NSFAS students drove the growth in this gap in 2020. The largest growth in the credits passed gap (column 3) occurred in the top quartile, where the gap grew from -2.2 credits (not statistically different from 0) to a statistically significant -9.7 credits. The data suggests this was driven by slight improvements among students not funded by NSFAS in quartile 4, with credits passed declining for NSFAS-funded students in this quartile. Together, these results suggest that, on average, non-NSFAS students in the top quartiles increased their credit loads, and also passed more credits, while top-performing NSFAS students responded to pandemic conditions by reducing their credit loads. This translated into a growth in the credits passed to credits taken achievement gap of 1.5 percentage points (column 5).

However, it appears as if different enrolment strategies in 2020 for NSFAS and non-NSFAS students may have muted gaps in performance. When considering only those who remain enrolled, NSFAS-funded students in quartile 1 took fewer credits than their non-NSFAS peers in both periods.[87] The increase in the credits passed gap when dropouts are excluded is similarly most stark at the bottom of the performance distribution in both COVID and pre-COVID periods. This is in line with higher dropout rates among non-NSFAS students.[88] Part of the dichotomy presented here is thus that wider negative achievement gaps between NSFAS-funded students and those not funded by NSFAS (when dropouts are excluded) may owe to the fact that relatively fewer NSFAS-funded students drop out. In this regard, NSFAS funding could be viewed as providing students with an important safety net to remaining enrolled. Whether or not this will translate into graduation remains to be explored.

Overall, if I exclude students who dropout from the analysis, in 2020 the gap in the share of credits passed grew by 3.6 percentage points in quartile 1, by 0.9 percentage points in quartile 2, 0.3 percentage points in quartile 3, and by 0.9 percentage points in quartile 4 relative to pre-COVID gaps in the same quartile.

Lastly, Figure 3.2 presents marginal effects of COVID-19 on credits taken and passed by GPA quartile and faculty. It is apparent that across faculties the overall performance improvements in 2020 were, as expected, driven by those in the bottom quartile. In the Commerce faculty, students in quartile 1 passed roughly one course more (18 credits) on average compared to those in quartile 1 pre-COVID. This follows just a 0.5 credit decline in credits taken, which is not statistically different from zero. Similar trends are apparent in Health Sciences,[89] EBE

[87] Since dropout is typically higher among non-NSFAS students, their inclusion with zero credits taken in the sample likely made it appear as if non-NSFAS students were taking relatively fewer credits.

[88] The exclusion of zero credits for these students seems to indicate that the credits passed average among non-NSFAS students would be higher, thereby increasing the gap between NSFAS-funded and non-NSFAS students on average.

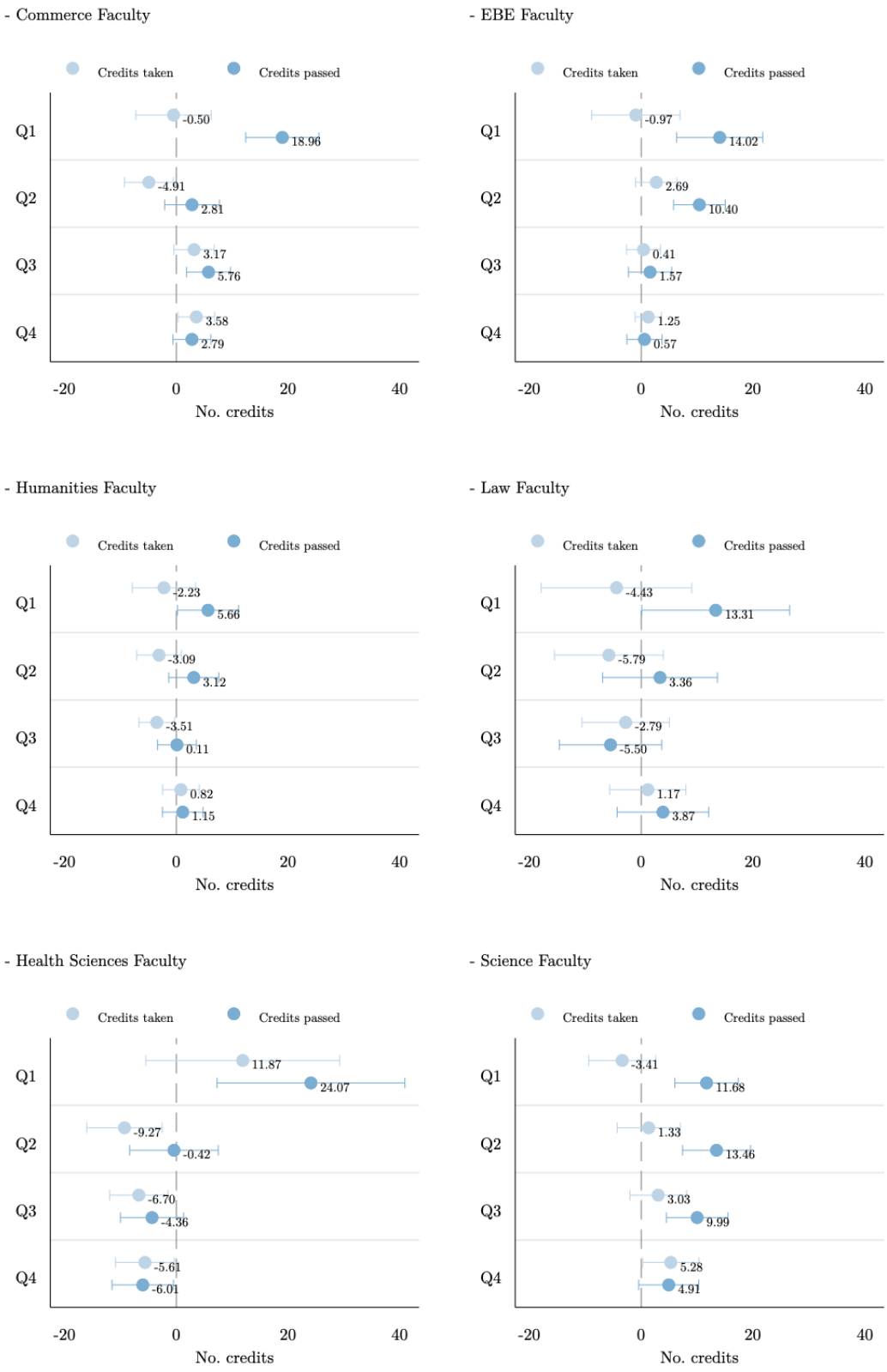
[89] The wider confidence intervals for the quartile 1 estimates in Health Sciences arise due to the small

(students in quartile 2 also experienced an increase), and the Science faculty. In the Sciences, all quartiles except the top quartile experienced significant increases in credits passed, this faculty thus displaying the most gains across its students. Improvements in quartile 1 were lowest among Humanities students, perhaps owing to marker and student familiarity with the continuous forms of assessment. Despite returning to campus earlier than students in other faculties, students in Health Sciences appear to have reduced their credit loads. This could be indicative of the burden of the pandemic on students involved with providing care during that time.

On the whole, results in this section are consistent with improved performance in 2020, especially among students with lower pre-pandemic GPAs. However, students who are NSFAS-funded, especially at the top of the performance distribution, seem to have fallen behind their non-NSFAS peers. Once I account for the fact that adjustments to enrolment in 2020 are different to previous years, the achievement gap at the bottom end of the distribution also becomes more stark.

sample of Health Science students who fall into the bottom quartile of the GPA distribution for a given cohort and year of study.

Figure 3.2 Marginal effects of COVID-19 on credits taken and passed by quartile of the previous year's GPA distribution and faculty



Source: Author's own calculations using UCT (2016-2020).
Notes: The sample comprises those in entrance cohorts 2016-2019, dropouts are included. Regressions include a triple interaction between the COVID effect indicator, quartiles of the previous year's GPA distribution, and an indicator variable for NSFAS funding. Regressions control for year of study.

3.6.2. Academic performance in 2021

In 2021, classes continued predominantly online at UCT, therefore performance may have continued to improve, it may have returned to pre-pandemic levels, or it may have worsened. This is assessed in Table 3.7, which provides the DD regression estimates for students' credits taken and passed in 2021, compared to students' pre-pandemic trends (excluding 2020 performance). The coefficients on the 2021 effect indicator show that in the year following the onset of the pandemic, trends in credits taken, passed, and the ratio of the two were below those prior to the pandemic. For example, the FE estimate on the share of credits passed shows that students passed 2.5 percentage points fewer credits in 2021, relative to pre-pandemic trends. With dropouts excluded, 2021 differentials become even more negative. Since dropout rates were substantially lower in 2021, the sample retained a greater share of students who were likely to be academically weaker.

Table 3.7 DD regression estimates, 2021 performance

	Additional credits taken			Additional credits passed			Credits passed to credits taken		
	OLS	FE	FE (excl. dropout)	OLS	FE	FE (excl. dropout)	OLS	FE	FE (excl. dropout)
2021 effect (P·T21)	2.677***	5.385***	-2.866***	-6.756***	-6.400***	-10.303***	-0.029***	-0.025***	-0.052***
	(0.882)	(0.809)	(0.666)	(1.037)	(0.898)	(0.866)	(0.006)	(0.005)	(0.004)
Post cohort (p)	3.251***	-	-	2.875***	-	-	0.000	-	-
	(0.582)	-	-	(0.754)	-	-	(0.004)	-	-
2nd year	-3.605***	-3.190***	6.237***	-3.469***	-3.134***	1.419***	-0.061***	-0.060***	-0.031***
	(0.424)	(0.406)	(0.264)	(0.435)	(0.398)	(0.376)	(0.003)	(0.002)	(0.002)
3rd year	5.841***	-1.828***	10.162***	7.674***	-0.299	5.696***	-0.021***	-0.068***	-0.030***
	(0.585)	(0.541)	(0.385)	(0.627)	(0.547)	(0.511)	(0.004)	(0.003)	(0.003)
4th year	-8.044***	-17.612***	-2.342***	-3.393***	-10.069***	-2.471***	-0.016***	-0.063***	-0.010**
	(0.884)	(0.917)	(0.745)	(1.017)	(0.910)	(0.871)	(0.006)	(0.005)	(0.004)
Constant	139.410***	143.280***	144.393***	120.126***	123.928***	127.138***	0.838***	0.855***	0.876***
	(0.374)	(0.221)	(0.152)	(0.489)	(0.235)	(0.216)	(0.003)	(0.001)	(0.001)
R^2	0.01	0.02	0.04	0.01	0.02	0.02	0.01	0.04	0.03
Observations	44 681	44 681	42 479	44 681	44 681	42 479	44 679	44 679	42 477

Source: Author's own calculations using UCT (2016-2021).
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses and are robust to the presence of repeated observations on the same individual. The sample comprises those in entrance cohorts 2016-2019. In the OLS regressions, the post cohort coefficient shows the differential for the 2016 and 2017 cohorts, the cohorts not affected in 2021 in their third and fourth years of study. Note that only students in years 3 and 4 are 'treated' in 2021.

Table 3.8, which shows the marginal effects of COVID-19 on the share of credits passed by GPA quartile and NSFAS status, shows that a greater share of academically weaker students remaining in 2021 is indeed likely to have been the case. Before dropouts are excluded, students not funded by NSFAS in quartile 1 continued to pass a significantly higher share of their credits in 2021 relative to their peers not funded by NSFAS pre-2020 (3.7 percentage points). When dropouts are excluded, however, it becomes apparent that non-NSFAS students who remained enrolled in 2021 actually performed worse relative to non-NSFAS students who remained enrolled

prior to 2020 (their share of credits passed is lower by 5.9 percentage points).

Table 3.8 Marginal effects of COVID-19 on the ratio of credits passed to credits taken by GPA quartile and NSFAS status, 2021 performance

	Ratio[a]	+ Controls	Excluding dropout
Students not funded by NSFAS			
Quartile 1	0.037**	0.040**	-0.059***
	(0.017)	(0.018)	(0.016)
Quartile 2	-0.019*	-0.004	-0.039***
	(0.011)	(0.011)	(0.010)
Quartile 3	-0.037***	-0.030***	-0.047***
	(0.009)	(0.009)	(0.008)
Quartile 4	-0.009	-0.010*	-0.023***
	(0.006)	(0.006)	(0.004)
NSFAS-funded students			
Quartile 1	-0.047***	-0.054***	-0.123***
	(0.017)	(0.017)	(0.016)
Quartile 2	-0.067***	-0.063***	-0.068***
	(0.015)	(0.015)	(0.014)
Quartile 3	-0.068***	-0.067***	-0.056***
	(0.014)	(0.014)	(0.013)
Quartile 4	-0.057***	-0.060***	-0.060***
	(0.015)	(0.015)	(0.014)
Observations	28 653	25 470	26 788

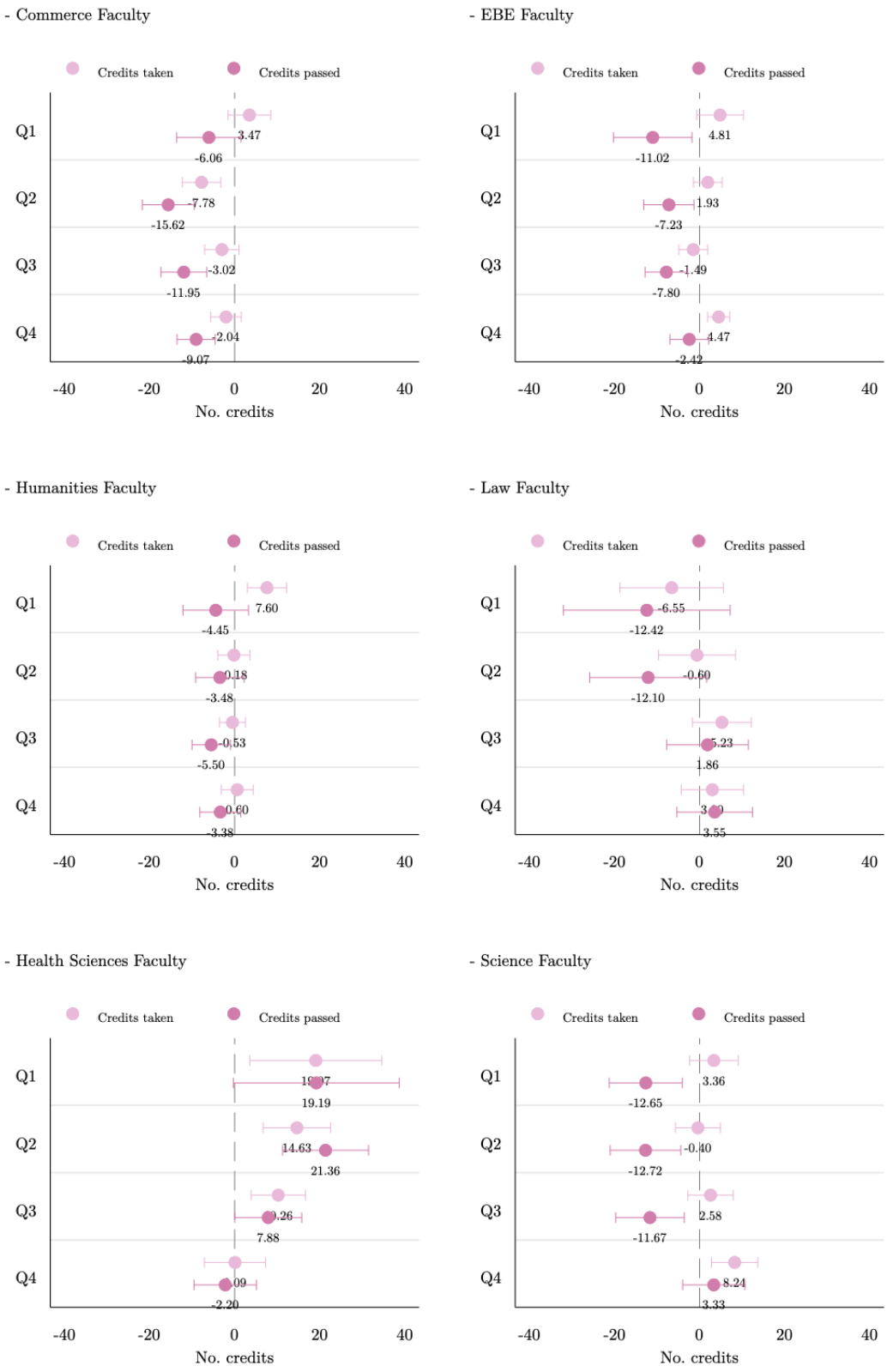
Source: Author's own calculations using UCT (2016-2021).
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses and are robust to the presence of repeated observations on the same individual. The sample comprises those in entrance cohorts 2016-2019. The regression with controls includes covariates for English first language, matric aggregate, and gender. The sample size reduces due to missing information for matric aggregate. First-year information is excluded due to lack of prior year GPA quartile.
[a] Full regression results are presented in Table B.2 in Appendix B.

Table 3.8 further shows that in 2021, NSFAS-funded students were performing significantly worse than their NSFAS-funded peers pre-2020, even before dropouts are excluded. NSFAS-funded students in quartile 1 passed 4.7 fewer credits than NSFAS-funded students prior to the pandemic, with those in the remaining three quartiles also passing a significantly lower share of their credits taken in 2021 by approximately 6-7 percentage points.

These patterns of general performance declines are shown by faculty in Figure 3.3. Health Sciences appears to be the only faculty in which students made credit gains. In all quartiles except quartile 4, credit loads increased relative to pre-pandemic trends, and students passed more credits too. This is in stark contrast to the performance of students in the Science faculty. Although the majority of science students made gains in credits passed in 2020, these students (with the exception of those in quartile 4) performed significantly worse in terms of credits passed in 2021 relative to pre-pandemic trends, despite no statistically significant adjustments in credits taken. Where content in subsequent years builds on foundational knowledge in programs offered in the sciences (e.g. maths, physics etc.), and potentially less content was taught in 2020, declines such as these could be expected. Given that Health Science students were treated quite differently during the pandemic (they returned to campus sooner, and were also dealing

with health provision itself during this time), it is not surprising that their performance reflects a different trajectory to the other faculties. In general, the type of assessment may differ by faculty, and this may determine the amount of leniency that could take place. For example, accounting or mathematics assessments may have less room for marker leniency than essay or oral based assessments.

Figure 3.3 Marginal effects of COVID-19 on credits taken and passed by quartile of the previous year's GPA distribution and faculty, 2021 performance.



Source: Author's own calculations using UCT (2016-2021).
Notes: The sample comprises those in entrance cohorts 2016-2019. Regressions control for NSFAS status and year of study. Dropouts are excluded.

Table 3.9 shows that performance declines in 2021 are accompanied by further growth in achievement gaps between NSFAS-funded students and those not funded by NSFAS. Comparing the bottom panel and top panel in column 1, it appears that the credits taken gap among students in the top quartile continued to grow despite both NSFAS-funded students and those not funded by NSFAS taking relatively more credits in 2021 relative to pre-pandemic trends (not shown here). This suggests that those not funded by NSFAS increased their credit load by a greater extent than NSFAS-funded students.

Table 3.9 Differential in credit measures between NSFAS students and students not funded by NSFAS (achievement gap), 2021 performance

	Credits taken		Credits passed		Credits passed to credits taken	
	(1) Taken[a]	(2) Excl. dropout	(3) Passed[a]	(4) Excl. dropout	(5) Ratio[a]	(6) Excl. dropout
Pre-COVID gap						
Quartile1	5.789***	-2.099*	-3.001	-9.901***	-0.013	-0.061***
	(1.901)	(1.236)	(1.983)	(1.861)	(0.013)	(0.012)
Quartile 2	-1.028	-5.381***	-8.813***	-12.750***	-0.033***	-0.058***
	(1.315)	(1.072)	(1.751)	(1.613)	(0.010)	(0.008)
Quartile 3	-1.524	-4.337***	-6.089***	-8.779***	-0.016**	-0.033***
	(1.176)	(1.002)	(1.579)	(1.447)	(0.008)	(0.007)
Quartile 4	-2.172	-4.410***	-3.645*	-5.896***	0.002	-0.012**
Post-COVID gap						
Quartile 1	-0.347	-5.564***	-16.220***	-20.369***	-0.097***	-0.125***
	(2.534)	(1.915)	(3.119)	(3.113)	(0.020)	(0.019)
Quartile 2	-2.887	-3.771**	-14.536***	-15.468***	-0.081***	-0.087***
	(2.048)	(1.831)	(2.782)	(2.718)	(0.015)	(0.015)
Quartile 3	-9.701***	-9.025***	-14.072***	-13.519***	-0.047***	-0.042***
	(2.060)	(1.839)	(2.666)	(2.562)	(0.014)	(0.013)
Quartile 4	-5.763**	-6.416***	-12.436***	-13.037***	-0.046***	-0.049***
	(2.323)	(2.246)	(3.043)	(2.996)	(0.013)	(0.013)
Observations	28 713	26 847	28 713	26 847	28 711	26 845

Source: Author's own calculations using UCT (2016-2021).
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses and are robust to the presence of repeated observations on the same individual. The sample comprises those in entrance cohorts 2016-2019. The table shows the marginal effects of NSFAS status on credit measures by GPA quartile and COVID-19 cohort. A negative effect indicates NSFAS-funded students are taking/passing fewer credits than their peers not funded by NSFAS. Regressions control only for year of study. [a] Full regression results are presented in Table B.2 in Appendix B.

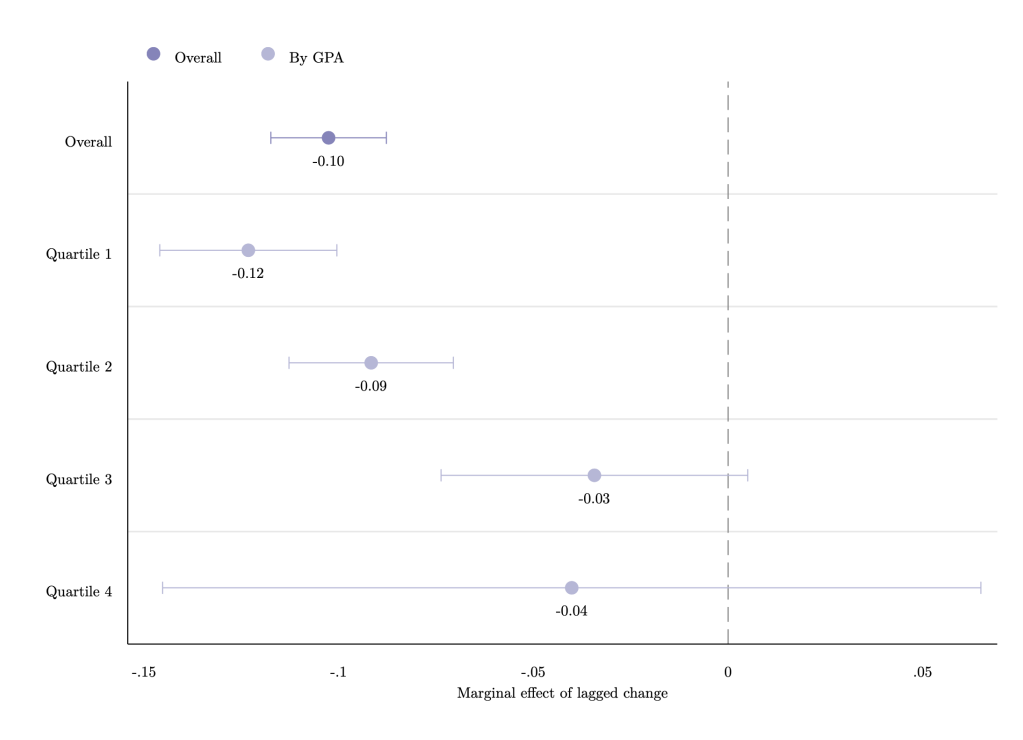
In terms of credits passed (column 3), gaps widened most substantially at the bottom and top of the distribution, as was the case in 2020. This appears to have been driven by NSFAS-funded students in all quartiles passing fewer credits in 2021 compared to pre-pandemic trends, with non-NSFAS students passing more. This translated into significant differences in the share of credits passed by NSFAS-funded students and those not funded by NSFAS in all quartiles in 2021 (column 5), with the gap in each quartile having grown by a minimum of 4 percentage points.

Differentially lower dropout in 2021, with a relatively greater share of non-NSFAS students dropping out, once again means that achievement gaps are wider when students who drop out are excluded compared to the case when dropouts are accounted for. Although lower dropout

seems positive, to the extent that there will likely be ripple effects and strong path dependence in performance emanating from the decline in 2021 performance into 2022, widening gaps is a concerning finding if post-school education is to disrupt rather than contribute to cycles of inequality.

Given the performance improvements observed in 2020, these declines in 2021 measures relative to pre-2020 trends likely imply an even greater downward trend in performance between 2020 and 2021. In this regard, it is evident that better performance does not necessarily imply learning gains (nor does it disprove loss of learning). To explore this further, I examine how potentially inflated performance in 2020 was related to performance in 2021, and how the strength of this association differs for students at different points of the GPA distribution. To this effect, Figure 3.4 shows the average marginal effect of the year 2021 on an indicator for improved performance in the year prior, by prior GPA quartile.

Figure 3.4 Marginal effects of lagged increase on ratio of credits passed to credits taken by quartile of the prior GPA distribution and year



Source: Author's own calculations using UCT (2016-2021).
Notes: The sample comprises those in entrance cohorts 2016-2019. Regressions exclude 2020. Confidence bands reflect 90% confidence intervals. Regressions control for year of study, and dropouts are excluded. Full regression results are presented in Table B.3 in Appendix B.

Specifically, the figure illustrates that the association between an increase in the share of credits passed the previous year is weaker in 2021 relative to pre-pandemic trends, by approximately 10 percentage points on average. This is true for all quartiles, but the size of the effect diminished the most for students in quartile 4 (a decrease of 12 percentage points). This shows a decline in the importance of prior performance in 2021. Together with the fact

that performance gains made in 2020 did not translate into sustained gains in 2021 (rather, there was an apparent lack of learning progress), the findings in Figure 3.4 suggests that performance improvements were likely artificial in the sense that gains in 2020 were not as strongly associated with passing a greater share of credits in 2021.

3.7. Concluding remarks

In this chapter, I set out to document the effects of university closures on undergraduate students' academic performance at the University of Cape Town. Three main concerns were raised that motivated the importance of understanding potential changes in students' academic trajectories. The first was that spuriously improved or even constant academic performance during the pandemic could unintentionally impact academic performance in the long term if students proceed to higher levels without sufficient baseline knowledge and competencies. Second, and relatedly, were concerns about the quality of grades and graduation as labour market signals of ability and knowledge acquired. A third concern was the potential for achievement gaps between students of different socioeconomic standing to have shifted, owing to the fact that household inequalities were likely to impact more directly on students' learning and performance.

I find significant improvements in credits passed and the share of credits passed to credits taken among students in 2020. This holds despite – or is perhaps even a result of – the fact that students seemed to adjust their credit load rather than their enrolment in response to the pandemic. For example, lowering credit loads may have enabled students to pass a greater share of their credits. Nonetheless, the decline observed in 2021 relative to pre-pandemic performance outcomes, together with the overall weaker correlation between performance in 2020 and performance in 2021, suggests that the credit improvements of 2020 were unlikely to represent true learning gains. Consequently, marks in 2020, as usual signals learning achievement, may be less informative of learning acquired. This means that performance and graduation are likely to be weaker as labour market signals of ability and knowledge acquired. A more thorough investigation of graduate employability will be required to ascertain whether or not the COVID cohorts of graduates face different employment outcomes.

The declines in credit-based measures of performance in 2021 suggest that there may have been a lag in the negative impact of the pandemic as evident in the student performance measures I have access to. However, if I account for the fact that a relatively greater share of academically weaker students remain in 2021, due to substantially lower dropout rates between 2020 and 2021, the magnitude of the pandemic effects on performance are smaller, although still negative.

For institutions grappling with student success, and for a sector seeking to dismantle its role in perpetuating cycles of inequality, recognising where heterogeneity in learning progress has shifted is important. Results suggest that students who were academically strong, and who were living in well-resourced households, were not as greatly impacted by COVID-19 closures as those students who were performing at the bottom end of the GPA distribution prior to the pandemic,

or who came from lower-income socioeconomic backgrounds (as proxied by NSFAS receipt). In this regard, results suggest an increased relevance of household inequalities on performance differentials since the COVID-19 pandemic.

Without intervention that recognises that not only could the negative effects of the pandemic continue to play out over time, but that the composition of the enrolled student body has changed as a result of the pandemic, gaps may continue to widen. Continued measurement of student performance will therefore be important, especially as institutions consider the role of online learning in a post-COVID era. Moreover, it will be fundamental to consider the challenges presented by COVID-19 not just as once-off occurrences, but rather in an on-going way. In this regard, should shocks to the system materialise in the future, both institutions and students will be able to respond more resiliently.

Of course, these results speak to the academic performance of students at one of the country's better-resourced institutions. The extent to which COVID-19 impacted the sector, and potentially shifted inequalities between institutions as well as within institutions, remains to be explored. Future research into the effects of the timing of students' return to campus would be an important addition to understanding the role of household inequalities on student performance.

Chapter 4
Studying to support? Exploring graduates' responsibilities to remit

Abstract

A large majority of Black South Africans remain restricted by intergenerational education and economic disadvantages. Labour market returns to tertiary qualifications in South Africa are high, but the share of Black youth accessing post-secondary education remains low. In such contexts, graduates may face responsibilities to financially support extended family networks. If graduate status induces responsibilities to offer support over and above income or employment status, graduates may experience a particular manifestation of (what is colloquially termed) 'black tax'.

Leveraging nationally representative data from South Africa, we fit a double hurdle model for remittance decisions. Results indicate that graduates are more likely to be remitters than other individuals, and that part of this responsibility arises from graduate status alone – that is, over and above labour market characteristics and living arrangements. We additionally observe a weaker relationship between remitting propensity and graduates' income as compared to other remitters; consistent with graduates facing a responsibility to remit, regardless of income earned. Conditional on remitting, however, graduates do not remit more than other remitters. Findings on graduate realities have implications for how tertiary education is conceptualised as a private vs. social benefit, which, in turn, has implications for tertiary education funding policy and the understanding of the intergenerational transmission of social inequality more generally.

4.1. Background, objectives, and contributions

Owing largely to the persisting ramifications of Apartheid, a large portion of South Africa's Black[90] population remains restricted by intergenerational education and economic disadvantages. Only 10% of Black individuals have a diploma, bachelor's or post-graduate degree,[91] and poverty is a daily threat for 76% of South Africans. This risk is greatest for the Black population (The World Bank, 2018).

Moreover, wealth accumulation for the Black population was curtailed by Apartheid policies, meaning that (generational) wealth inequality in South Africa is stark (Chatterjee et al., 2022; Mbewe & Woolard, 2016; The World Bank, 2018). Although there exist a variety of state transfers for social support,[92] access to formal structures of private financial support (e.g. bank loans) are not available to those without asset surety. Networks offering informal, private financial support thus remain important. Public and media discourse has colloquially termed these financial contributions to family 'black tax' (Daya & Mpete, 2017; Mhlongo, 2019).[93]

Labour market returns to tertiary qualifications in South Africa are high (Branson et al., 2013; Keswell & Poswell, 2004; Salisbury, 2016), but a low share of the population has access to these returns. Black tax is thus relevant to Black graduates in particular, since the completion of post-secondary education may come with responsibilities to meet family and household needs. This may be a result of increased ability to offer financial support, or for example, reciprocity for investments in graduates' education in the past (du Toit & Neves, 2006). Where graduate status further induces responsibilities to offer support over and above income or employment status, graduates may experience a particular manifestation of black tax.

While I acknowledge that black tax is not unique to graduates,[94] I am particularly interested in how it manifests for this group of individuals for two core reasons. First, although returns to tertiary education in South Africa are large and positive, without adequately understanding behaviour with respect to private financial support, and particularly how this manifests among

[90] Black African. I.e. my sample does not include Indian and Coloured individuals, although in South Africa these populations are sometimes classified as 'Black'. I focus on Black individuals given accounts of the differential responsibilities Black students face to meet familial obligations and repay debt (see e.g. Webb, 2021).

[91] See Siyaphambili website here. Diplomas, bachelor's degrees and post-graduate degrees awarded at universities, colleges or other post-school institutions are typically viewed as high value qualifications. These contrast to all qualifications which additionally include short-course qualifications or certificates.

[92] Often, state transfers are a source of private transfers. For example, access to state transfers (especially the old age pension) plays an important role in ensuring private safety nets for the unemployed (Klasen & Woolard, 2008).

[93] Generally, black tax broadly refers to financial contributions that Black professionals make to less fortunate family members. See Mhlongo (2019) for a more in-depth discussion.

[94] Nor may it be unique to Black African individuals. However, addressing the extent which graduates of other population groups experience manifestations of black tax is beyond the scope of this paper.

higher income earners, one may inadvertently make inaccurate conclusions about the income equalising and wealth generating capacity of tertiary education completion. Thus despite the fact that private transfers can work to lower the persistence of poverty among receiving households (Posel, 2016), where they hamper individuals' own wealth accumulation (for example), they can manifest as an intergenerational transmission of social inequality, and maintain or replicate horizontal inequalities (see e.g. Oppel, 2022).

Second, interrogating graduate responsibilities[95] is relevant to tertiary education funding policy. If tertiary education were to be financed through a graduate tax, for example, Black graduates' incomes could be seen to be taxed twice – privately by transfers sent to support family members and publicly via the graduate tax. Moreover, a key motivation for a cost-sharing model of funding is that there is a substantial private return to post-schooling. My results, however, show that a share of this private benefit may be re-conceptualised as a social benefit when accounting for black tax.

This study thus contributes to the growing body of research on support practices and patterns in South and Southern Africa through the lens of black tax (e.g. Fongwa, 2019; Mangoma & Wilson-Prangley, 2019; Oppel, 2021; Webb, 2021). To the best of my knowledge, this study is the first of its kind to quantitatively analyse private transfers within the context of black tax and the responsibilities associated with the completion of post-secondary education in South Africa. Using the latest available evidence from a nationally representative household survey, we re-examine inter-household, private transfer (remittance hereafter) behaviour in South Africa. Detailed information on remitters collected at the individual level allows us to interrogate who within a household unit is sending remittances: individuals with varying interests or preferences may respond differently to the needs of certain family members compared to others, or may face differing responsibilities to offer financial support.

My first step is to understand the intra-household allocation of income to determine how households make remittance decisions. That is whether remittance expenditure is allocated from a pool of household income, or whether it is allocated from the resources under the control of the remitter – and according to their preferences or responsibilities. For example, if households pool income, then the presence of a graduate in the household may affect the value of remittances sent whether the graduate is the remitter or not. This is because a graduate's labour market income is likely to be, on average, higher, which would increase the pool of household income from which remittance expenditure is drawn. However, if income pooling is rejected, the effect of a graduate's income on remittances may only matter if they are the remitter. Following this, my second step is then to look at the determinants of remitting.

[95] I use the term graduate responsibilities to refer to potential expectations on graduates to make financial contributions to their families. Responsibilities may arise from feelings of obligation, reciprocity, altruism or indebtedness, among others. Given that tax typically has a negative association, I prefer the term 'graduate responsibility' to 'black tax' as it encompasses the various motivations for sharing.

How I conceptualise the household is important for the purpose of my analysis. Households in the data comprise both non-resident members (who live under the household's roof for at least 15 days a year) and resident members (who usually sleep in the household for at least four nights a week). For the analysis, I define a household to include resident members only.[96] Reference to remitters' or graduates' households is always with regard to their household of residence – which may be their original household (e.g. childhood household), or a new household. A remittance is a private transfer sent[97] to any individual who is not resident in the remitter's household (hence it is an inter-household transfer). Questions on remittances are asked of all adult respondents, not only of labour migrants.

I aim to answer following questions:

1. Does an individual's income, and whether or not the individual is a graduate, play a role in determining remittance sending, and the value of the remittance they send, over and above the collective resources available in the household? That is, are remittances sent from a common pool of household income, or does selection into remitting and the amount of the transfer depend on a remitter's individual income?

   If the income-pooling hypothesis is rejected, this would suggest that an individual's characteristics and preferences would likely influence their decision – or responsibility – to remit. I then ask:

2. Who sends transfers and to whom? How does this relate to the structure of a remitter's household of residence, and does it differ by the remitter's graduate status?[98]

The remainder of the chapter is set out as follows: in Section 4.2, I provide background on black tax, private transfers, and remittances in South Africa. This is followed by a brief review of the responsibilities that graduates may face to extend financial support to members of their family. In Section 4.3, I introduce theory on intra-household decision making, and in Section 4.4 present the empirical specification used to test the income-pooling hypothesis, followed by a discussion on the double hurdle model used in the multivariate analysis. In Section 4.5, I present my data. In Section 4.6 the results of income-pooling hypothesis tests are discussed, along with descriptive statistics, the results of the double hurdle model, and findings on intra-household transfers. I offer concluding remarks in Section 4.7.

[96] This narrower definition of the household can be used since household income comprises the sum of resident members' incomes only.

[97] This analysis does not consider remittance receiving from the point of view of individuals or their households.

[98] A graduate of post-secondary education, or not.

4.2. Related literature

4.2.1. Private transfers and remittances

Networks offering informal, private financial support are important in South Africa. There exist a variety of state transfers for social support,[99] but access to formal structures of private financial support (e.g. bank loans) are not available to those without asset surety. Wealth accumulation for the Black population was curtailed by apartheid policies, meaning that (generational) wealth inequality in South Africa is stark (Chatterjee et al., 2022; Mbewe & Woolard, 2016; The World Bank, 2018). Chiteji and Hamilton (2002) study wealth accumulation in the United States and find that differing levels of relative economic status within a family (e.g. having parents or siblings that are poor) explain a non-trivial portion of the wealth accumulation gap. This likely owes to private transfers from relatively better off to relatively worse off family members.

Public and media discourse has colloquially termed these financial contributions to family 'black tax'. While private support networks in South Africa are fairly well researched,[100] literature considering financial transfers through the lens of black tax is less established. A recent study by Mangoma and Wilson-Prangley (2019) focusses on understanding the lived experiences of those making financial transfers. Fongwa (2019) contemplates black tax as a higher education public good and Oppel (2021) links racial and structural inequality in Namibia to the concept of black tax in a post-apartheid context. A recently published book of essays on the topic (Mhlongo, 2019) queries whether black tax is a 'burden or *ubuntu*[101]' by delving into the complexities of the everyday lived experiences of the book's contributors and their kin. The stories exemplify private networks of financial support in South Africa, and while contributors' opinions vary about whether black tax is a burden or *ubuntu*, none dispute its existence.

Writing about support networks in Namibia, Oppel (2021) describes black tax as "neither an in-depth account of ethnic tradition nor one of Namibia's political economy alone, but a narrative that stresses a mutual constitution between them." In this vein, Mzobe (2019) deftly captures the notion that black tax was borne of apartheid, but is perpetuated by the less obvious, but equally damaging, ramifications of that era that persist in my society today. While the migrant labour system established during apartheid was a central driver of remittance sending in the past[102]

[99] Often, state transfers are a source of private transfers. For example, access to state transfers (especially the old age pension) plays an important role in ensuring private safety nets for the unemployed (Klasen & Woolard, 2008).

[100] E.g. di Falco and Bulte (2011), du Toit and Neves (2006), Jensen (2003), Klasen and Woolard (2008), Maitra and Ray (2003), Posel (2001a, 2001b, 2016), and Sagner and Mtati (1999).

[101] The word *ubuntu*, common to many indigenous languages and cultures in South Africa, literally translates as 'humanity', but encompasses the idea that 'I am because you are'. It largely speaks to concern for others and/or community wellbeing, which can in turn govern solidarity, mutual exchange, and reciprocity. It is against this culture of humanity that opponents of the term black tax advocate for terms such as 'collective family responsibility' instead (e.g. Maqetuka, 2019).

[102] Black migrants were prevented from settling permanently in urban areas. Thus remittances were

(Wilson, 2011), various authors have suggested that prevailing conditions of unemployment and poverty are key drivers of current private transfers and remittances (Jensen, 2003; Magubane, 2016; Sagner & Mtati, 1999).

At an individual and household level, there are a variety of factors that may affect the decision to remit and the value of transfers sent. Remitting may be influenced by altruism, bargaining (Cox & Fafchamps, 2008), as well as obligations or responsibilities to fulfil implicit contracts.[103] Additionally, transfers may be induced by specific ties at the household the remitter is sending to, or by receipt of government grants such as the child support grant received for non-coresident children, for example.

On the other hand, others have argued that remittance sending may be reduced by public transfers that crowd out private transfers. Jensen (2003) provides evidence that public pensions crowd out private transfers, although, Maitra and Ray (2003) find this is the case only for poor households, not non-poor households. Where individuals establish a household or ties to a household away from their original household, ties to the new household can crowd out remittances. In particular, children are likely to compete for household or remitter resources (Posel, 2001b).

Age, or life-cycle considerations, may impact remittances, especially where older migrants may expect to return permanently to their households of origin in the near future. Posel (2001a) also suggests that expectations of reciprocity drove remittance sending in South Africa in the 1990s. It is, however, hard to elicit the true preferences behind remitting. Posel (2001b) finds evidence that migrants with more precarious forms of employment remit more (in absolute terms). Part of these preferences are likely strong willingness to remain connected to the household of origin, operating through incentives to secure a slice of 'the pie' in households of origin. Thus, self-interested motives can also impact sharing, over and above income earned. On the other hand, I suggest that if those in better paid or more secure forms of employment face greater responsibilities or obligations to remit, the opposite result to that observed by Posel (2001b) may be observed.

4.2.2. Graduate responsibilities

The literature on the relationship between education acquisition and offering familial support is fairly well established. This literature particularly suggests that parents pay for their children's education in the form of a loan, that is later repaid through support offered, especially during old age (Bollard et al., 2011; Frankenberg et al., 2002; Lillard & Willis, 1997; Nyathi, 2019). Investment in children's education may also form part of a household survival strategy – du

not only an important means of economic support for households of origin, but also a means for the migrant to retain ties with the household. See Wilson (2011) for how South Africa's migrant labour system generated and entrenched poverty and racial inequality.

[103] The black tax narrative particularly speaks to the responsibility shouldered by graduates and working professionals in Black communities (Magubane, 2016; Mhlongo, 2019).

Toit and Neves (2006) suggest education provides access to urban economic opportunity, but with this comes the expectation that employed adult children will later support the household. Complementing this, qualitative findings from a study in Cameroon highlight parents' feelings that once educated, graduates should provide for the needs of their families (Mbah, 2014).

Du Toit and Neves (2006) further evidence scenarios in which individuals and households continually invest in children's education, even when it seems to be an irrational strategy (e.g. the student is failing, or qualifications have not led to jobs in the past), because the expected payoff is anticipated to be so large. Such notions could fuel feelings of indebtedness (see also Chauke, 2019) on behalf of students to support their families once their studies are complete. This is echoed by Ghanaian university graduates interviewed for Caldwell's (1965) early work on the topic.

Evidence on the relationship between education and remittances (more specifically) is typically found in international migration literature, but is suggestive of a positive relationship especially for the value of remittances sent (Aristei, 2013; Bettin et al., 2012; Bollard et al., 2011). However, Bollard et al. (2011) provide a number of reasons why the effect of education on remittances may be ambiguous. For example, more highly educated migrants earn more, which enables them to remit more. On the other hand, those who are more highly educated may come from more well-off families, minimising the need to send remittances. They may be more likely to migrate with their family and/or have less intention of returning home, both of which reduce the role of remittances in retaining ties with home communities. Alternatively, where remittances serve as repayment for past investments in education, this again increases the probability of remitting for this group of migrants. In the context of the analysis in this chapter, where remitters are not necessarily migrants only, education may additionally increase the likelihood that an individual sets up their own household, which may induce inter-household transfers observed as remittances in the data.

4.3. Theoretical framework

4.3.1. Household decision making

Becker's (1974) unitary model commonly provides a starting point for theoretical work in the field of household decision making. The model assumes a benevolent head, who allocates pooled household income among all members. It is assumed that the household's utility function is identical to that of the head, since the head, through their concern for the welfare of other members,[104] fully incorporates all members' utility functions. Since Becker's (1974) pioneering work, however, scholarship on household decision making has grown extensively, featuring arguably more realistic 'collective' models of decision making. The general framework for collective models differs from the unitary model framework in that intra-household allocations

[104] I.e. the head is not dictatorial (Becker, 1974).

can be influenced by differing preferences[105] and characteristics of household members. Adapting the framework in Bourguignon and Chiappori (1992) and Bourguignon et al. (2009), I present a general framework of household behaviour in the following scenario:

Assume a two-person household with members A and B who have utility functions u^A and u^B respectively. The household can consume n private goods and N public goods. Thus, private consumption bundles, x, for members $m = A, B$ are represented by $\boldsymbol{x}^m = (x_1^m, \ldots, x_n^m)$ and the household consumption of public goods can be written as $\boldsymbol{X} = (X_1, \ldots, X_N)$. Each person has utility represented by $u^m = (\boldsymbol{x}^A, \boldsymbol{x}^B, \boldsymbol{X}, \boldsymbol{a})$, where $\boldsymbol{a}$ is a vector of individual characteristics that may determine preference factors (e.g. age, race, education) that affect the utility gained from consuming different commodities. Note that this general formulation does not exclude the possibility that members care about each other's utility (the existence of altruism, for example, as in Becker's original model).

The overall budget constraint can be represented as $y = \boldsymbol{p} \cdot (\boldsymbol{x}^A + \boldsymbol{x}^B) + \boldsymbol{P} \cdot \boldsymbol{X}$ where $\boldsymbol{p}$ is the price vector for private goods, $\boldsymbol{P}$ the price vector for public goods, and y the total household income (equal to $y^A + y^B$ when each member's individual income can be independently observed). A utility maximising household's demand for private goods can be expressed as $x^* = \xi(y, \boldsymbol{a}, \boldsymbol{z})$, where $\boldsymbol{z}$ is a vector of what the literature terms 'distribution factors'. These affect the relative bargaining power of members (and as such decision processes) without directly affecting preferences or the household budget constraint (e.g. relative incomes, relative wages, the control of land (Bourguignon et al., 2009)).

Under a unitary model, it is assumed that the household's utility function is identical to that of the head (Becker, 1974). The model thus assumes that the household maximises a single utility function of household preferences and is analysed as if it is a single decision-making unit.[106] Under this framework, the household utility function should depend on income, y, and preference factors, $\boldsymbol{a}$, but not on distribution factors, $\boldsymbol{z}$. This implies $\frac{\delta \xi_i(y, \boldsymbol{a}, \boldsymbol{z})}{\delta z_k} = 0$ for every commodity, i,[107] and distribution factor, k. The unitary model thus implies that household demand is unaffected by individual income (or any other factor that does not affect household preferences).

Under a collective model, members jointly determine how resources are accessed and redistributed via cooperative 'sharing rules'[108] (Bourguignon et al., 1993; Bourguignon & Chiappori, 1992; Chiappori, 1992) or bargaining processes which can be either co-operative

[105] Which are potentially altruistic (Chiappori, 1992).

[106] There exists a utility function $U(\boldsymbol{x}^A, \boldsymbol{x}^B, \boldsymbol{X}, \boldsymbol{a})$, such that for every $(y, \boldsymbol{a}, \boldsymbol{z})$, the vector $(\boldsymbol{x}^A, \boldsymbol{x}^B, \boldsymbol{X})$ maximises $U(\cdot)$ subject to the budget constraint.

[107] Although not a commodity in the traditional sense, remittances still fit in to this framework since they represent an expenditure decision.

[108] I.e. if member A gets $\theta(\boldsymbol{p}, y)$, member B gets $y - \theta(\boldsymbol{p}, y)$. The authors do not provide details on how this rule comes to be, but assume the resulting allocations are efficient (i.e. each individual's utility function is maximised to the effective budget of each – Deaton (2018) provides a summary).

(e.g. Manser & Brown, 1980) or non-cooperative (e.g. Lundberg & Pollak, 1993).[109] Regarding bargaining processes, distribution factors, $\boldsymbol{z}$, affect bargaining power, which in turn affects household decision making.

4.3.2. The income-pooling hypothesis

A common approach to test for a unitary model is to test an income-pooling hypothesis. This is because under a unitary model, only total income – and not income composition – matters. Testing the income-pooling hypothesis is thus equivalent to testing:

$$\frac{\delta \xi_i(y, \boldsymbol{a}, \boldsymbol{z})}{\delta z_k} = 0, \tag{4.1}$$

where in my case the distribution factor, $\boldsymbol{z}$, is individual's relative income.[110] While a rejection of income-pooling falsifies a unitary approach, it does not necessarily support the collective approach, or importantly, a particular model thereof (Thomas, 1990). It is thus important to note that the analysis in the chapter does not attempt to identify a particular collective model nor recover the underlying decision-making or bargaining processes. However, if remittance-sending households do not behave as a unit, then I have scope to examine differences in the individual remittance behaviour of graduates compared to other household respondents.

In related research, Browning et al. (1994) find that among Canadian couples with no children, increasing an individual's relative share of household income affects final allocations. In South Africa, Ambler (2016) shows that altering household income shares can shift bargaining power: Black women who experience an increase in income when they become pension eligible are more likely to be primary decision makers in their households, compared to non-eligible women.

Further evidence against unitary decision-making models in South African households and extended families can be found in Bertrand et al. (2003), Burger and Kreuser (2017), Duflo (2003), Gummerson and Schneider (2013), and Posel (2001b) and Quisumbing and Maluccio (2003). Thomas (1996), however, shows that there may be variation across race groups in South Africa.

[109] Haddad and Kanbur (1990) provide a summary of cooperative and non-cooperative bargaining frameworks.

[110] Note that relative income is unlikely to be exogenous with respect to unobserved preferences (time-allocation preferences) and thus its validity as a distribution factor in the test of equation 4.1 will not be as strong as completely exogenous measures such as divorce laws or gender of welfare recipients (Haddad & Kanbur, 1990). Other distribution factors that studies have used which are less likely than income to be correlated with unobserved preferences include members' relative education, unearned incomes, and family background factors (Burger & Kreuser, 2017). To address the potential endogeneity of relative income, I test the robustness of my results by examining the whether the share of total household asset value owned by residents is significant, once total income available to the household is controlled for. Results are substantively very similar. However, the ownership of assets may still be correlated with labour market outcomes and thus be endogenous. Results should be interpreted with this in mind, and addressing the potential endogeneity, especially of income, bears further analysis in the future.

Investigating the effect of parental income on children's education levels, he provides evidence in support of income pooling in Black and White households, but rejects the income-pooling hypothesis, and hence the unitary model, for Coloured and Asian households.

It is important to note here, that although evidence of income pooling is typically understood as evidence in favour of the unitary model, Thomas's (1996) findings do not preclude a collective model of household decision making in Black and White households. Although income is typically not pooled in collective models (e.g. Bourguignon & Chiappori, 1992), Manser and Brown (1980) present a bargaining model in which individuals do pool their resources. Final allocations, though, are determined via a bargaining rule which accounts for differences in individuals' utility functions. The limited extent to which an income-pooling hypothesis test can provide evidence in favour or against a specific model of household behaviour serves to emphasise the challenge that scholars have faced not only in deriving models of household decision making, but also in identifying them empirically.

In her earlier work, Posel (2001b) rejects the income-pooling hypothesis for remittance expenditure based on a sample of households from KwaZulu-Natal province surveyed relatively soon after the end of apartheid. Authors have suggested that in recent years there has been shrinkage in the extent of kinship ties in South Africa (Harper & Seekings, 2010), as well as changing patterns of mobility and migration (Hall, 2016). While remittances were traditionally sent from men in urban areas to their rural households of origin, Posel et al. (2006) find that the expansion of the state old age pension has promoted women's migration. Transfers may also be governed less by altruism or self-interest but by feelings of reciprocity or responsibility. This, especially with the expansion of post-school qualifications among Black youth in recent years.[111] Where remitters and their households have changing characteristics and formations over time, patterns of decision making similarly may not be static and therefore warrant re-examination.

4.4. Empirical framework

4.4.1. Testing the income-pooling hypothesis

To test the income-pooling hypothesis for remittance expenditure, I adapt and expand on the empirical framework laid out in Posel (2001b). I estimate the following equation:[112]

$$R_{ih} = \gamma_1 Y_h + \gamma_2 Y_i^L + \gamma_3 Y_i^O + \boldsymbol{X_h'}\theta + \boldsymbol{X_i'}\delta + \epsilon_{ih}, \tag{4.2}$$

where R_{ih} in equation 4.2 is either an indicator for whether individual i in household h reports sending a positive remittance or not, or in a separate case, it takes the positive value of cash and in-kind remittances sent by a remitter in the last month. Y_i^L is an individual's monthly income from the labour market; Y_i^O is their monthly income from other, non-labour market sources

[111] See Siyaphambili website here.

[112] Note I also expand on the theoretical framework above by allowing for households of more than two individuals, which may be non-nuclear and span more than two generations.

(other income hereafter); and Y_h is household monthly income excluding individual i's income. I separate income in labour market income and other income to account for the fact that it may be used differently depending on its source.

$\boldsymbol{X_h}$ is a vector of household-level controls including the number of children and location; $\boldsymbol{X_i}$ is a vector of individual-level regressors including gender, a quadratic in age, education, and employment status. ϵ_{ih} is the individual-specific error and therefore includes all unobserved factors that affect the amount of remittance sent, assumed to be uncorrelated with all included regressors. Under the income-pooling hypothesis $\gamma_1 = \gamma_2 + \gamma_3$. Income pooling is rejected if $\gamma_1 \neq \gamma_2 + \gamma_3$. This equates to testing equation 4.1 with $z = Y_i^L + Y_i^O$.

4.4.2. Determinants of remitting

If the income-pooling hypothesis is rejected, there is motivation to further consider how an individual's personal characteristics may affect their decision to remit, as well as their choice of how much to remit. Given that individuals may choose not to remit at all, there is a corner solution response where remittance value is zero. The Tobit model allows estimation where the dependent variable includes corner solution responses, but it assumes that the underlying mechanisms, in this case, of the decision to remit and the decision of how much to remit, are the same.

To allow for the fact that individuals' characteristics can have a differing effect on participation in remitting and the amount sent, one can implement a more flexible 'double hurdle' approach (Cragg, 1971) to model remittance decisions. Furthermore, the double hurdle model allows for double censoring: the fact that zeros may arise because individuals choose not to remit, or because potential remitters are not able to. Regarding the latter point, circumstances such as financial constraints or fixed costs might dictate a zero outcome (Bettin et al., 2012). For example, remitting may create an expectation of continued remitting in the future. In the international literature, double hurdle models have been widely used to model expenditure and consumption decisions, for example, cigarette consumption (Jones, 1989), household demand for clothing and married women's labour supply (Blundell & Meghir, 1987), consumption of prepared meals (Newman et al., 2003), education expenditure (Mussa, 2013), and remittance decisions (Bettin et al., 2012; Sinning, 2011).[113]

Following Wooldridge (2010), the double hurdle model (in the context of remittance spending) can be specified as follows: Assume remittance value, R, is generated as $R = s \cdot w^*$, where s is a binary variable that determines whether R is zero or strictly positive, and w^* is a

[113]Note that the hurdle model differs from sample selection models. In selection models, the outcome is not always fully observed – e.g. zero wages indicate that the potential wage, if an individual were to work, is unobserved. In the hurdle model, possible outcomes include true zero values, so there is no sample selection problem (Wooldridge, 2010). Since I am interested in actual rather than potential expenditures in this study, I thus model the probability of a response at zero, as well as the density for positive values.

continuously distributed, non-negative latent variable. Both s and w^* are functions of observable and unobservable characteristics, but w^* is only observed when $s = 1$, in which case $R = w^*$. A key assumption of the model is conditional independence of s and w^*. This implies that the mechanisms determining s and w^* are independent of each other. Then, s follows a probit model,

$$P(s = 1|x) = \Phi(\boldsymbol{x}\boldsymbol{\gamma}), \tag{4.3}$$

and $w^* = \boldsymbol{x}\boldsymbol{\beta} + u$ follows a truncated normal distribution. The density of R given the covariates and $R > 0$ is

$$f(R|\boldsymbol{x}, R > 0) = [\Phi(\frac{\boldsymbol{x}\boldsymbol{\beta}}{\sigma})]^{-1}\phi[\frac{R - \boldsymbol{x}\boldsymbol{\beta}}{\sigma}]/\sigma, \quad R > 0, \tag{4.4}$$

where σ is the variance and the term $[\Phi(\frac{\boldsymbol{x}\boldsymbol{\beta}}{\sigma})]^{-1}$ ensures the density integrates to one over positive values. The conditional independence assumption means the parameters in equation 4.4 can vary from those in 4.3. The truncated normal hurdle model is thus estimated in two independent steps – by a probit regression of s in the first step and a truncated regression of R in the second. Alternatively, if w^* is assumed to follow a lognormal distribution, then $w^* = exp(\boldsymbol{x}\boldsymbol{\beta}+u)$. In this case, the model is termed the lognormal hurdle model, and the second step involves an Ordinary Least Squares [OLS] regression on $log(R)$.

Specification tests

I use a robust (adjusted) Wald statistic to test whether a Tobit model or a double hurdle model is more appropriate. The choice of a Wald test over the more commonly used Likelihood Ratio [LR] test in this setting owes to the fact that the LR test is not valid for models fit with probability weights or cluster-robust standard errors. Given that this study draws on survey data, accounting for these sample design features cannot be avoided.

Since the hurdle model nests the Tobit, the Wald test equates to estimating a truncated normal hurdle model and testing the restrictions imposed by the Tobit model, namely $\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$. This equates to testing restrictions of the form $\hat{\gamma}_j = \hat{\beta}_j/\hat{\sigma}$ (Wooldridge, 2010). The robust Wald statistic is an adjustment of the Wald statistic, W, as $(d - k + 1)W/(kd)$, where k is the number of restrictions imposed, and d is the total number of sampled Primary Sampling Units [PSUs] minus the total number of strata. Under the null hypothesis that the restrictions hold, $(d - k + 1)F/(kd) \sim F(k, d - k + 1)$.

If the Tobit is rejected, it is relevant to test whether the lognormal hurdle model or the truncated normal hurdle model is closer to the true model in the underlying population (Rahmani & Wooldridge, 2019). Vuong (1989) develops a test that compares non-nested models by measuring the closeness of each to the true model. This approach, however, relies on log-likelihoods which, again, are not valid when estimation accounts for survey data design. Rahmani and Wooldridge (2019) extend Vuong's test to survey data analysis, and I implement this version of the test here. Under the null hypothesis, both models are equally close to the true

model – that is, the models evaluated at the pseudo-true values[114] fit equally well on average. The alternative is that one of them fits better. Note that because the probit (participation) equation is the same in both the truncated normal hurdle model and the lognormal hurdle model, I need only compare the pseudo-log-likelihoods from the truncated regression and the OLS regression on $log(R)$ (the amount equations). Testing equates to calculating the difference in pseudo-log-likelihoods for each observation, i, and running a regression of the difference on a constant. The test statistic is the t-statistic on the constant, and the sign of the constant determines which model fits better.[115]

In my specification, the Tobit model imposes 20 restrictions of the form $\hat{\gamma}_j = \hat{\beta}_j/\hat{\sigma}$. A robust Wald test of these restrictions[116] rejects the Tobit model in favour of the truncated normal hurdle model at the 1% significance level.[117] Results of the model fit test following Rahmani and Wooldridge (2019) reject the lognormal hurdle model in favour of the truncated normal hurdle model at the 10% significance level ($p = 0.057$). I present results of the Tobit and lognormal hurdle models in Appendix C.

The participation equation

I begin by estimating the following probit regression, which estimates the probability of an individual sending a positive remittance, conditional on individual and household characteristics:

$$s_{ih} = \gamma_1 Y_i^L + \gamma_2 Y_i^L \cdot g_i + \gamma_3 Y_i^O + \gamma_4 Y_i^O \cdot g_i + \boldsymbol{X_h'}\gamma + \boldsymbol{X_i'}\gamma + \epsilon_{ih}, \tag{4.5}$$

where s_{ih} is an indicator equal to 1 if an individual reports sending a positive remittance value. Again, $\boldsymbol{X_h}$ is a vector of household-level controls including the number of children and location; $\boldsymbol{X_i}$ is a vector of individual-level regressors including a male dummy, a quadratic in age, education, and employment status. I further introduce an interaction between the income variables and graduate status, g_i, to test whether there are differential effects for income received by graduates versus individuals without a post-secondary education (other individuals hereafter). For example, if the majority of non-labour market income for other individuals comes from child support grants received for non-coresident children, this may more readily induce remitting than graduates' income from non-labour market sources (which may be more likely to be returns from financial assets, for example).

[114] Values computed from estimation accounting for survey design.

[115] This will depend on the way the difference is constructed. If the difference is computed as the pseudo-log-likelihood of the truncated regression less the pseudo-log-likelihood of the OLS regression on $log(R)$, then a positive and statistically significant difference would reject the OLS regression in favour of the truncated regression. See Rahmani and Wooldridge (2019) for more detail.

[116] Following the estimation of the truncated regression I test whether the β_j coefficients are equal to a constant defined as $\hat{\gamma_j} \times \hat{\sigma}$.

[117] $W = 44.24; p = 0.003$.

The amount equation

Thereafter, I explore the determinants of remittance value conditional on remittances being sent. I estimate the following:

$$R_{ih} = \beta_1 Y_i^L + \beta_2 Y_i^L \cdot g_i + \beta_3 Y_i^O + \beta_4 Y_i^O \cdot g_i + \boldsymbol{X_h'}\beta + \boldsymbol{X_i'}\beta + \epsilon_{ih}, \tag{4.6}$$

where R_{ih} is the amount of cash and in-kind remittances sent in the last month. I again include an interaction between graduate status, g_i, and the income variables to test for a differential effect of graduate income on the amount remitted. All variables across equations 4.5 and 4.6 are the same and defined as previously noted.

4.4.3. Measurement error and endogeneity

A limitation of the income-pooling test specification in equation 4.2 is that if income is measured with error, coefficients in the income-pooling hypothesis test will be biased towards zero. Income captured via a survey is vulnerable to measurement error because respondents may be reluctant to disclose income, may understate or overstate their income, or may have trouble accurately recalling income (especially before-tax income), and, in household surveys where a household responds on behalf of individual members, the respondent does not have full information for all household members.

In the survey data I use, individuals are administered individual questionnaires and therefore report their income from each source directly, ameliorating the latter measurement error concern for the individual income measures. For the test of the income pooling hypothesis, my main concern regarding measurement error is thus that total household income, reported by the household head, will be reported with more error than individual income. If this is the case, the coefficient on household income β_1 would be biased towards 0. If this is the case, the result of testing the relative significance of individual income compared to household income may be driven by measurement error in household income. In households with more than one remitter, I can control for measurement error in household income by estimating a household fixed-effects [FE] model that replaces y_h with a fixed effect μ_h:

$$R_{ih} = \mu_h + \gamma_1 Y_i^L + \gamma_1 Y_i^O + \boldsymbol{X_h'}\theta + \boldsymbol{X_i'}\delta + \epsilon_{ih}. \tag{4.7}$$

If households behave as a unit, an individual's income should have no additional impact on the amount sent. That is, γ_1 and/or γ_2 in equation 4.7 should not be statistically different from zero.

Where household- or individual-level unobservables are relevant for selection into remitting and/or remittance value, and are correlated with the other covariates, coefficients from the probit and truncated regressions in equations 4.5 and 4.6 will be biased. Additionally, where individuals who want to remit more work more, the income variable will be endogenous. It is beyond the scope of this chapter to address these issues, but existing literature has used methods such as

instrumental variable estimation to overcome endogeneity of income (e.g Bettin et al., 2012). Results should be interpreted with these cautions in mind.

4.5. Data and definitions

The above regressions are fit to data from Wave 5 of the National Income Dynamics Study [NIDS] (Southern Africa Labour and Development Research Unit [SALDRU], 2018), the first nationally representative panel study in South Africa. In Wave 5 the sample was topped up to account for high attrition in high-income areas (Branson & Wittenberg, 2019). The Wave 5 sample includes 37 368 successfully interviewed respondents.

An important contribution of the NIDS data is that it facilitates research of intra-family allocations and remittance behaviours, particularly by offering a wealth of information on cash and in-kind transfers made and received, as well as on income received at both the individual and household level.

For the purpose of this analysis, a graduate is defined as an individual who has reported successfully completing a post-secondary education qualification. Students who complete their final three years of secondary school in the general education stream[118] are eligible to obtain a school leaving qualification called the National Senior Certificate [NSC] or 'matric'.[119] Students who complete an NSC with the appropriate pass level – a bachelor's pass – can go on to complete a post-secondary qualification at a university. Those completing matric with an NSC or diploma pass can enrol for post-secondary education at a TVET college, and universities of technology also accept students with a diploma pass into certain courses. Students who do not complete the NSC still have the option to complete a post-school[120] certificate or diploma that does not require an NSC qualification, but these qualifications do not fall under post-*secondary* education by my classification.

The sample of interest is adult, Black respondents and their households in NIDS Wave 5. Questions about cash and in-kind contributions sent to non-residents[121] are asked of all household residents aged 15 and above as part of the individually-administered adult questionnaire. A remitter is defined as an individual who reports having sent a positive contribution in the last month.[122] A household is classified as a remittance-sending household if at least one resident

[118] This refers to an academic stream which prepares students for further studies in a university or college. It contrasts with a technical or vocational stream.

[119] The NSC is commonly termed 'matric' in South Africa. I use the terms interchangeably.

[120] Post-school education refers to any education that takes place after this compulsory schooling, outside the general education stream. Compulsory schooling in South Africa occurs from the year in which a child turns seven until Grade 9 or the age of fifteen, whichever occurs first (South African Schools Act, 1996).

[121] I.e. remittances can be sent to non-resident household members, or to individuals who are not part of the household at all.

[122] This measure has a higher response rate than the annual measure and is less likely affected by recall bias. Moreover, income measures are reported as monthly amounts.

member reports making a cash and/or in-kind contribution in the last 12 months. A household is considered Black if the household head is Black.[123]

4.6. Results and discussion

4.6.1. Remittance-sending households

Before testing the income-pooling hypothesis, it is useful to contextualise the average characteristics of remittance-sending households. Black households comprise 76% of households successfully surveyed in Wave 5. Remittance-sending Black households represent 20.5% of all Black households.[124] This is in line with national estimates for Black households in Posel (2001a), but is lower than rates reported in surveys from other developing countries such as Indonesia, where 88% of households report sending a transfer (LaFave & Thomas, 2017).[125] The low share of remittance sending in South Africa is perhaps unsurprising, though, given extensive social support programmes (Jensen, 2003).

Table 4.1 presents the mean characteristics of Black households that report sending remittances compared to those that do not. All characteristics presented differ significantly between these two household types. The majority of households reporting remitters report only one resident remitter (80%), however, this is partly a result of 54% of remittance-sending households being single-person households (i.e. the remitter is the sole resident).[126] While the majority (64%) of Black households are situated in urban localities, the share of households in urban areas is predictably higher for remittance-sending households (73%).

Remittance-sending households report fewer resident children on average, and the share of households with at least one child in each of the age groups is significantly lower in remittance-sending households than in other households. As suggested by (Posel, 2001b), other household residents – and especially children – may crowd-out remittance sending. This is one of the potential reasons that I observe fewer residents, and specifically fewer children, residing in households that do not send remittances. Another reason may owe to the endogeneity of household formation. Where remittances are sent to support non-coresident biological children, one would observe a higher frequency of remittances sent from households in which remitters do not live with their children. There is thus a selection mechanism, which is related to household formation, affecting observed transfers.

Remitter households further differ on a number of dimensions. They are smaller, have older

[123] Where no head is reported, the modal race of household residents is used. Only four households report do not report information on a head.

[124] The overlap between remittance-sending and remittance-receiving households is low: just 19% of remittance-sending households also report receiving transfers from a non-resident or non-member (not shown here).

[125] Similarly in Malaysia, 31% of households report receiving a remittance (Vogel & Korinek, 2012) compared to around 20% of households receiving remittances in my data.

[126] There may, however, be non-resident members.

Table 4.1 Characteristics of all Black households compared to remittance-sending Black households

	All	Remittance-sending household?	
	households	No	Yes
Household monthly expenditure	5 690.91	5 121.38	6 664.32***
	(8 711.22)	(7 770.61)	(9 171.35)
Household monthly income	8 489.11	7 125.92	11 278.59***
	(24 639.18)	(27 049.68)	(13 955.9)
Household has a resident graduate (yes=1)[a]	0.26	0.23	0.34***
Household head is male (yes=1)	0.52	0.46	0.67***
Average age of residents	30.49	29.8	32.28***
Household has at least 1 child aged 7 to 17 (yes=1)	0.37	0.43	0.22***
Household has at least 1 child under age 7 (yes=1)	0.31	0.36	0.19***
Number of children 7 to 17	0.64	0.75	0.36***
Number of children under 7	0.45	0.53	0.27***
Number of residents	3.03	3.29	2.38***
Household is a single person household	0.39	0.33	0.54***
Household has only 1 remitter (yes=1)[b]	0.22	0.00	0.80***
Urban location (yes=1)	0.64	0.61	0.73***
Household has at least one pensioner (yes=1)	0.16	0.19	0.09***
Household receives at least one state grant (yes=1)	0.42	0.49	0.25***
Observations	8 490	6 705	1 723

Source: Author's own calculations using NIDS Wave 5.
Notes: Data are weighted using post-stratification weights. There are 8 428 households that have valid responses to remittances sent. Statistically significant differences in the characteristics of households that send remittances and households that do not are indicated by *** p<0.01, ** p<0.05, * p<0.1. Standard deviations are in brackets for select variables. Income and expenditure values are in Rands.
[a] This figure is conditional on residents in the households having non-missing education status. There are 1 718 remittance-sending households with non-missing graduate status, and 6 699 households that do not send remittances which have non-missing graduate status. Thus N=8 479.
[b] Conditional on residents in the household having valid responses to the question on contributions given. N=8 428.

members, are more likely to be male headed, and are more likely to have a graduate resident. They also have significantly higher household income and expenditure. Having a graduate in the household who is accessing labour market returns to a university or college degree will contribute to the significantly higher income levels observed in remittance-sending households. If income is pooled, then the presence of a graduate in the household may have an effect on remittances sent irrespective of whether the graduate is the remitter or not. However, if income-pooling is rejected, then the presence of graduates who do not remit would not necessarily affect the amount remitted. Thirty-four percent of remittance-sending households have a resident graduate, compared to only 23% of other households. This is partly driven by the higher prevalence of remitting reported by graduates. Graduates comprise just over one quarter of individuals who report sending remittances — roughly double the share of graduates found in the population.

4.6.2. Testing the income-pooling hypothesis

The first step in understanding how best to measure whether there are potential differences in expectations of support placed on individuals of differing characteristics (e.g. graduates

versus other individuals) is to check whether the income-pooling hypothesis holds. If it holds, I would expect household characteristics rather than individual characteristics to determine whether a household is a remittance-sending household. To this extent, I am interested in whether remittance expenditure is drawn from a common pool of household income, or whether a remitter's income matters more in this regard. In the former case, the coefficients on other total household income and individual income should be similar. In the latter case, the coefficient on individual income should be statistically different. A rejection of the income-pooling hypothesis would suggest that the resources a remitter potentially brings into the household may be allocated differently from other resources – that is, according to the remitter's preferences or responsibilities.

Table 4.2 presents the results of regressions to test the income-pooling hypothesis. Columns 1 and 2 show results for the probability of a remittance being sent, and columns 3 and 4 show results for remittance value, conditional on a positive value being sent.

Table 4.2 Regressions testing the income-pooling hypothesis

	Probability of remittance sent		Amount remitted	
	(1) LPM	(2) FE	(3) OLS	(4) FE
Other household income	-0.000	-	-2.243	-
	(0.000)	-	(5.036)	-
Individual labour market income	0.004***	0.004**	50.688***	55.932***
	(0.002)	(0.002)	(14.597)	(22.751)
Individual income from non-labour sources	0.000	0.000	30.704	5.642
	(0.000)	(0.000)	(334.179)	(57.887)
Controls	Yes	Yes	Yes	Yes
Observations	18 966	18 966	1 664	1 664
R-squared	0.18	0.12	0.17	0.05
Number of households	-	8 458	-	1 528
F-test: individual labour market income + non-labour market income = other household income (p-value)	0.00	-	0.03	-

Source: Author's own calculations using NIDS Wave 5.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses and are robust to clustering at the cluster level. Data are weighted using post-stratification weights. The dependent variable in columns 1 and 2 is a binary variable = 1 if an individual is a remitter, and in columns 3 and 4 it is value of cash and in-kind remittances sent in the last month. Additional controls include gender, a quadratic in age, categories of education (no schooling, primary, incomplete secondary, matric, qualification not requiring matric, and post-secondary), dummy variables for whether the individual is in wage employment and/or casual/self-employment respectively, household location (rural/urban), the number of children under 6 in the household, the number of children aged 6-16, and indicator variables for whether each of the mother and father are co-resident or not. The income variables are scaled by 1 000, to represent the effect of a R1 000 change in income on the dependent variable. Results are robust to the exclusion of single-person households.

The sample in the probability of remitting specification comprises 18 966 Black individuals who responded to whether they made a cash or in-kind contribution to someone not co-resident in their household in the last 12 months (this includes maintenance and child support payments)[127]

[127] Fifty-nine individuals who have missing education information are excluded, along with 27 individuals missing information across the other covariates.

The sample in the remittance value specification (columns 3 and 4) consists of 1 664 remitters who responded with a positive value to the question on how much was sent in the last month – both cash and in-kind. Some remitters report zero incomes (5% of the estimation sample). I thus enter both income and the value of remittances sent as levels rather than logarithms. Columns 1 and 3 show the resulting coefficients on income from the OLS regression[128] and columns 2 and 4 show the results of a FE specification, which verifies whether results are robust to measurement error in the household income variable.

The results in Table 4.2 reject the income-pooling hypothesis for both selection into remitting and the amount remitted. F-tests in columns 1 and 3 show that equality of individual income and household income is rejected at the 1% and 5% levels respectively.[129] Results in column 2 confirm that the rejection of the income-pooling hypothesis for selection into remitting is robust to measurement error in the household income variable. Results in column 4 show that the coefficient on labour market income increases in size and remains significant, indicating that the income-pooling hypothesis is also rejected for the value of remittances sent. These results suggest that households do not operate as a unit with respect to remittance decisions. As such, I now turn consider which personal characteristics of individuals may affect selection into remitting.

4.6.3. Determinants of remittance sending: descriptive analysis

Table 4.3 contrasts the characteristics of individuals who remit compared to those who do not remit. On average, remitters fair better in the labour market than those who do not. Remitters are more likely to be in wage employment, less likely to receive a social grant and are more likely to be graduates compared to those who do not remit (13% of individuals who do not remit are graduates compared to 29% of remitters). In other words, those who remit are distinct on observable characteristics. The higher prevalence of graduates among remitters provides suggestive evidence that graduates may be disproportionately affected by black tax.

Graduates who do not remit earn more than other remitters (remitters without a post-secondary qualification) on average, suggesting that higher income alone is not a sole determinant of remittance sending. However, both graduate remitters and other remitters' income comprises a relatively greater share of total household income than does the income of individuals who do not remit. This may represent a measure of the bargaining power that these individuals may leverage within their households to shift spending in line with their preferences.

Given the relatively small sample of graduates, and the potential for the income distribution

[128] I confirm the linear probability model [LPM] results with a probit model. Results are substantively similar and are available on request. When substituting the share of total household assets owned by a resident in place of the individual income measures, the income-pooling hypothesis is rejected in all specifications, except in the household fixed effect estimation for the value sent. This owes to insufficient within variation for this measure in households with more than one remitter.

[129] Income values are as reported in the survey. When income is deflated to a base month, coefficients do not change.

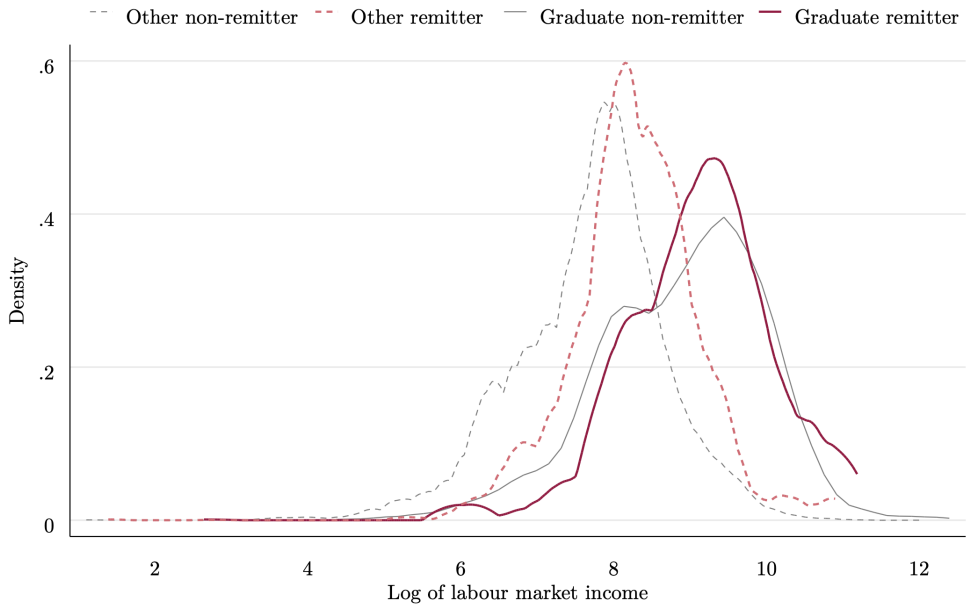
Table 4.3 Characteristics of those who remit and those who do not by their graduate status

	Remitter?		Remitter=no Graduate?		Remitter=yes Graduate?	
	No	Yes	No	Yes	No	Yes
Highest level of education						
No schooling	0.06	0.02	0.06	0.00***	0.03	0.00***
Primary schooling	0.14	0.08	0.16	0.00***	0.11	0.00***
Incomplete secondary schooling	0.44	0.30	0.50	0.00***	0.42	0.00***
Matric	0.17	0.20	0.2	0.00***	0.28	0.00***
Qualification not requiring matric	0.06	0.12	0.07	0.00***	0.17	0.00***
Post-secondary (post-matric)	0.13	0.29	0.00	1.00	0.00	1.00
Age (years)	35.89	36.80	36.04	34.92***	37.02	36.28
Male (yes=1)	0.44	0.66	0.44	0.40**	0.71	0.54***
Mother's education (years)[a]	5.35	5.25	5.03	7.49***	4.35	7.45***
Father's education (years)[a]	4.75	4.59	4.41	6.99***	3.81	6.45***
Highest level of mother's education is post-secondary[a]	0.06	0.06	0.04	0.14***	0.02	0.14***
Highest level of father's education is post-secondary[a]	0.05	0.05	0.03	0.13***	0.02	0.14***
Individual labour market income	2 024.78	6 991.56	1 133.07	8 223.69***	5 054.50	11 896.79***
	(5 852.18)	(8 975.01)	(2 793.96)	(13 235.52)	(6 985.35)	(11 313.60)
Individual income from non-labour sources	756.03	491.24	642.45	1 544.84	527.18	418.25
	(16 796.07)	(3 863.50)	(1 732.30)	(47 273.30)	(4 457.92)	(1 765.54)
Individual income as a shareof household income	0.36	0.75	0.34	0.55***	0.75	0.75
Wage employment (yes=1)	0.29	0.77	0.24	0.62***	0.73	0.85***
Casual/self employment (yes=1)[b]	0.10	0.14	0.10	0.08	0.13	0.14
Individual receives a government grant	0.36	0.15	0.38	0.19***	0.18	0.10**
Individual receives a child support grant	0.26	0.13	0.27	0.18***	0.14	0.09*
Observations[c]	17 375	1 677	15 664	1 662	1 212	455

Source: Author's own calculations using NIDS Wave 5.
Notes: Data are weighted using post-stratification weights. Statistically significant differences in the characteristics of graduates and other individuals are indicated by *** p<0.01, ** p<0.05 and * p<0.1. Standard deviations are in brackets for select variables.
[a] Parental education contains high rates of missingness, especially for father's education. The number of missing observations for maternal education is 76 for remitters and 1 043 for those who do not remit. The number missing paternal education is 252 and 3 018 for remitters and those who do not remit, respectively. Rates of missingness are slightly lower among graduates. Parental education variables are conditional on non-missing information.
[b] Three remitters did not respond to the question on casual/self employment. One is a graduate. Forty-four individuals who do not remit are missing information on casual/self employment (three are graduates) and 21 are missing information on wage employment (two are graduates). Summary statistics are conditional on non-missing information.
[c] Forty-nine individuals who do not remit are missing information on their graduate status. No remitters are missing this information. Thus the observations in the columns for those who do remit do not sum to the observations in the first columns.

of this group to be quite distinct from the other group, I examine kernel density estimates of logged labour market income by remitter and graduate status. Figure 4.1 shows that the distributions overlap a fair bit, illustrating that graduate remitters do not have incomes entirely out of range of the other respondents. This is important for identification in the regression results later in this section. The figure further shows that the distribution of remitters versus those who do not remit is more similar for the graduate group than the other group. The distribution of other remitters is more to the left than the distribution of other individuals who do not remit, which may suggest that income may be a more important determinant of who remits among the other group compared to the graduate group.

Figure 4.1 Kernel density functions of logged income by graduate and remitter status



Source: Author's own calculations using NIDS Wave 5 (post-stratified weights).

Table 4.3 additionally shows that the share of other remitters in wage employment and casual/self-employment[130] is higher than among graduates who do not remit, suggesting that employment may likely be a key driver of remittance sending too. However, employment and income are rough approximations of bargaining power, as they may be the result of time-allocation bargaining processes. Nonetheless, earning a greater share of household income, or having greater job security, may enhance individuals' bargaining power in their households.

Table 4.4 shows that remitters live in very differently structured households than those that do not remit. These household living arrangements will be endogenous to the decision to remit.

[130]I separate wage (or salaried) employment and casual/self-employment since job security and income stability is likely to differ between these two employment types, with the latter being a more precarious form of employment than the former. Casual employment is defined in the NIDS questionnaire as irregular and short-term employment.

Table 4.4 broadly shows an increased autonomy and reduced reliance on support structures associated with the nuclear family as consideration moves from those who do not remit, to those who do. Graduates within the group that do not remit are marginally less reliant than the other individuals, and then graduate remitters are even less reliant than other remitters.

Table 4.4 Characteristics of those who remit and those who do not by their graduate status

	Remitter?		Remitter=no Graduate		Remitter=yes Graduate	
	No	Yes	No	Yes	No	Yes
Mother is co-resident (yes=1)	0.26	0.05	0.27	0.21***	0.06	0.04**
Father is co-resident (yes=1)	0.10	0.02	0.10	0.09	0.02	0.02
At least 1 biological child co-resident (yes=1)[a]	0.45	0.23	0.44	0.49**	0.21	0.28**
At least 1 grandmother co-resident (yes=1)[b]	0.03	0.00	0.03	0.01***	0.00	0.00
At least 1 grandfather co-resident (yes=1)[b]	0.01	0.00	0.01	0.00	0.00	0.00
HH has at least 1 child aged 7–17 (yes=1)	0.60	0.21	0.61	0.48***	0.20	0.23
HH has at least 1 child under 7 (yes=1)	0.48	0.19	0.49	0.39***	0.19	0.18
Number of children aged 6–16 in the HH	1.16	0.35	1.21	0.78***	0.33	0.38
Number of children under 6 in the HH	0.79	0.26	0.82	0.57***	0.27	0.22
Household size	4.83	2.42	4.98	3.78***	2.41	2.34
Individual lives alone (yes=1)	0.15	0.52	0.14	0.26***	0.53	0.50
Household is in an urban location (yes=1)	0.57	0.73	0.55	0.73***	0.72	0.75
Observations	17 375	1 901	15 664	1 667	1 212	455

Source: Author's own calculations using NIDS Wave 5.
Notes: Data are weighted using post-stratification weights. Statistically significant differences in the characteristics of graduates and other individuals are indicated by *** p<0.01, ** p<0.05 and * p<0.1. HH abbreviates household.
[a] Respondents without children are assigned the value zero on this indicator. Note, male respondents are not asked whether they have any biological children in the survey. We can therefore only identify biological children with whom they reside.
[b] Note we can only link a grandchild to the grandparent via a personal identifier when a child's parent has been co-resident with the grandparent at least once during the course of the panel. In the case where a child has always lived with the grandparent but not their parent, this share may be underrepresented.

The biggest difference in household structure, however, is between those who remit and those who do not – that is, selection into remitting. Compared to those who do not remit, remitters are less likely to co-reside with a biological child, are more likely to live alone, or are more likely to reside in households with fewer children and other individuals in general. That is, the household structure of graduate remitters and other remitters is closer in similarity than the household structure of graduates and others who do not remit. For example, remitter households are of a similar size and equally likely to have children, whereas among those who do not remit, graduates' households are significantly smaller than other households and significantly less likely to have children, on average. Note that since men are not asked whether they have any biological children in the survey, the 'live with a biological child' indicator reflects the combined differences in the incidence of having a biological child and differences in co-residency patterns.

These findings suggest that household structure likely also affects participation in remitting, in addition to the labour market and education characteristics of individuals. Being a graduate may be correlated with an individual's level of autonomy – due in part to labour market factors – to start their own household. So rather than an increased responsibility (conditional on income)

alone, remitting may also be a function of residing in a different place (i.e. other individuals could be equally called upon to remit but are more likely to reside in their original household). This is explored further in the multivariate analysis.

Focusing now on differences between graduate remitters and other remitters, Table 4.3 shows that other remitters earn on average less than half the labour market income of graduates, are more likely to receive a social grant (the majority of which are child support grants) and are significantly less likely to be in wage employment. In fact, 5% of all remitters report no income from any source. This could owe to measurement error, or it could occur if an intra-household transfer was made between the remitter and another resident, which would enable the individual to send a remittance.

Table 4.3 further shows that remitters are, in general, more likely to be male. However, 46% of graduate remitters are women compared to only 29% of other remitters. This could be suggestive of a couple of different things. First, post-secondary education may increase the likelihood that women live and work away from their household of origin. Second, increased education could increase a women's economic freedom and hence ability to offer financial support or, perhaps, the expectation to offer financial support to those outside her home. As Posel (2001a) notes, historically, most migrants were men as they faced better urban economic opportunities and household decision-making hierarchies restricted women's mobility. Graduating with a post-secondary qualification could play an important part in shifting gender norms in this regard.

Graduate remitters are more likely to reside with a biological child compared to other remitters (Table 4.4), but again it is unclear whether this is a result of graduate remitters being more likely to have a biological child, more likely to reside with a biological child, or a combination of both. More individuals who do not remit live with their biological children than do remitters, a fact that is likely driving the high share of remittances sent to children (37%)[131] observed in Table 4.5 below. Unfortunately, though, unless all of a male respondent's children have been co-resident with them at some point in the panel, I do not know how many children a male in NIDS has. Thus, despite the fact that together Tables 4.3 and 4.5 indicate that a substantial number of remittances are likely sent from non-coresident fathers to their children, I do not have a complete birth register to explore this in more detail.

4.6.4. Determinants of remittance value: descriptive analysis

Table 4.5 shows that a similar share of graduate and other remitters also send contributions to their children (38%), but graduates are more likely to send contributions to their parents (39%) than other remitters (28%). This is consistent with lower parental co-residency rates for graduates shown in Table 4.4. Since other remitters are less likely to have wage employment (Table 4.3), this may contribute to a delay in the formation of their own household (see Klasen & Woolard, 2008). It is also consistent with the idea that parents may support their children's

[131] This includes foster, step, and adopted children.

Table 4.5 Characteristics of remitters by identity of recipient and remitter graduate status

	All	Others	Graduates
Remittance(s) sent to [...] (yes=1)[a]			
Children	0.37	0.38	0.38
Parents	0.32	0.28	0.39***
Grandparents	0.02	0.02	0.01*
Siblings	0.10	0.09	0.11
Other family members	0.16	0.17	0.14
Other non-family members[b]	0.09	0.10	0.06**
Remittance(s) sent to [relation to household] (yes=1)			
A non-resident household member	0.20	0.20	0.20
A non-household member	0.81	0.81	0.82
Number of people remitted to	1.11	1.08	1.16**
Value of remittances sent in the last month	1 575.56	1 419.99	1 983.52**
	(1 649.04)	(1 443.23)	(2 037.54)
Remittance as a share of total individual income[c]	0.53	0.63	0.30
Observations	1 677	1 212	455

Source: Author's own calculations using NIDS Wave 5.
Notes: Data are weighted using post-stratification weights. Statistically significant differences in the characteristics of graduate and other remitters are indicated by *** p<0.01, ** p<0.05 and * p<0.1. Standard deviations are in brackets for remittance values.
[a] Note the proportions in these categories do not sum to 100% as some remitters send to more than one person.
[b] This category includes boyfriends and girlfriends, lodgers or relatives of lodgers, and household help.
[c] This figure is conditional on positive income. Seventy-four other remitters report zero income while only two graduate remitters report zero income.

education in the form of a loan that is later repaid. Although, since I am not able to observe the purpose of cash remittances sent, it may be that remittances sent to parents are intended to be spent on children (where children co-reside with their grandparents, for example).

Note, that because of the way the NIDS survey is designed, I may still be observing remitters from within their household of origin (as termed in the migration literature). Questions on remittances are not only asked of labour migrants, but of all adult respondents. Thus, I am looking at remitters who may have moved out of their original households and are sending money home, as well as remitters who remain in their original households but who send money to relations living elsewhere, for example a non-resident member who has migrated to work or look for work. This could be the case for the one-fifth of remitters who send remittances to non-resident members of their household. However, I see that the majority of remitters send to individuals who are not household members at all (Table 4.5).

Table 4.5 shows that an average of R1 576 was remitted in the month prior to the survey, but graduates remit significantly more on average.[132] This is likely enabled by their higher average incomes observed in Table 4.3. The number of individuals whom a remitter supports is also slightly higher for graduates than for other remitters. This could suggest that graduates face

[132]It is important to note that remittances may fluctuate over the course of the year, and thus some variability in the value is likely to owe to this fact. Additionally, where individuals remit over the course of their lifecycle, cross-sectional estimates will underestimate the extent of transfers.

a greater responsibility to offer support than others, although a causal conclusion cannot be drawn from this data alone. Overall, remitters send roughly half their total monthly income, on average. Graduate remitters, however, spend 30% of their incomes compared to 63% among other remitters. The difference is not statistically significant.

Table 4.6 presents the average frequency and value of cash and in-kind remittances sent in the last month by the recipient's relationship to the sender.

Table 4.6 Average frequency and value of remittances by recipient's relationship to the remitter

	All	Share of remittances	Remittances sent to [...] by other individuals	Remittances sent to [...] by graduates
Frequency of cash and in-kind contributions sent to **children** in the last year	13.28	-	12.79	14.58
Value of cash and in-kind contributions sent to **children** in the last month	1 287.17	-	986.08	2 066.97***
Observations	708	36%	519	187
Frequency of cash and in-kind contributions sent to **parents** in the last year	11.61	-	11.82	11.19
Value of cash and in-kind contributions sent to **parents** in the last month	1 197.18	-	1 102.57	1 374.61
Observations	540	27%	363	171
Frequency of cash and in-kind contributions sent to **siblings** in the last year	8.73	-	8.80	8.64
Value of cash and in-kind contributions sent to **siblings** in the last month	850.25	-	828.27	888.95
Observations	199	10%	136	62
Frequency of cash and in-kind contributions sent to **grandparents** in the last year	9.38	-	9.28	10.09
Value of cash and in-kind contributions sent to **grandparents** in the last month	708.22	-	672.14	968.18
Observations	45	2%	36	9
Frequency of cash and in-kind contributions sent to **other family** in the last year	12.60	-	12.74	12.25
Value of cash and in-kind contributions sent to **other family** in the last month	2 437.36	-	2 491.72	2 339.83
Observations	276	14%	209	65
Frequency of cash and in-kind contributions sent to **other non-family** in the last year	10.11	-	10.07	10.52
Value of cash and in-kind contributions sent to **other non-family** in the last month	1 011.95	-	924.02	1 486.50**
Observations	218	11%	178	38

Source: Author's own calculations using NIDS Wave 5.
Notes: Data are weighted using post-stratification weights. Statistically significant differences in the remittances sent by graduate and other remitters are indicated by *** p<0.01, ** p<0.05 and * p<0.1. All values are in Rands. Where remitters have missing education information, the number of observations in the final two columns won't sum to the number of observations in the 'all' column. Note that each remittance does not equal one remitter. A remitter may send to multiple children, or to a child and a parent, for example.

The majority of remittances (36%) were sent to children in the month prior to the survey. Although roughly the same share of graduate and other remitters remit to children, graduate remitters remit significantly higher amounts to their children than other remitters. While Table 4.5 shows that a significantly lower share of graduate remitters support non-family members, Table 4.6 shows that among those that do, the amount sent is higher. The group that both graduates and other remitters send the highest remittances to, is other family. This is perhaps unsurprising since this category includes spouses.[133]

Very few transfers are directed towards grandparents, possibly reflecting the good coverage of the old age pension. As alluded to previously, the patterns of remittances will be affected by household formation and with whom the remitter resides, but the low share of grandparent co-residency Table (4.4) suggests that I am not simply missing transfers to the elderly because of co-residency arrangements.

Together, Tables 4.3 to 4.6 are suggestive of largely differing patterns of remittances sent by graduates and other individuals. Graduate remitters are, on average, more likely to support their parents and support more individuals more generally. Graduate remitters also send more on average, although it is not clear whether this can be attributed to higher levels of income, greater responsibilities, or an interplay of these factors. The regression analysis in the following section explores this further.

4.6.5. Results of the double hurdle model

The descriptive tables provide strong motivation to explore remitting in a multivariate setting. I first present and discuss results from the participation probit. Thereafter, I present the determinants of remittance value, conditional on a positive amount being sent.

The regression in column 1 of Table 4.7 panel (a) estimates the relationship between remitting and education, controlling only for gender and a quadratic in age. As suggested by the descriptive statistics, graduates of post-secondary education are more likely to be remitters. In this simple specification, the average marginal effects [AMEs] on all the education categories except no schooling are significant, indicating that those with more education are more likely to remit. The AME for post-school qualifications not requiring matric and post-secondary qualifications show a 8 and 14 percentage point increased probability of remitting relative to those with incomplete secondary education. Matriculants, are also more likely to remit than those with incomplete secondary education, but the AME is smaller at 5 percentage points. Finally, those with primary schooling are 3 percentage points less likely to remit than those with incomplete secondary.

The second regression (column 2), which introduces income and employment variables, suggests that the relationship between education and remittance sending is partly a function of improved labour market outcomes.

[133] Posel (2001b) finds evidence that spouses induce higher transfers.

Table 4.7 Probit average and conditional marginal effects for the probability of remitting

(a) Average marginal effects	(1)	(2)	(3)
Education (base: incomplete secondary)			
No schooling	-0.023	-0.014	-0.010
	(0.021)	(0.020)	(0.019)
Primary school	-0.026*	-0.015	-0.009
	(0.014)	(0.014)	(0.014)
Matric	0.047***	0.028**	0.009
	(0.017)	(0.014)	(0.012)
Qualification not requiring matric	0.083***	0.059***	0.041**
	(0.024)	(0.022)	(0.020)
Post-secondary	0.136***	0.054***	0.043***
	(0.019)	(0.017)	(0.016)
Age (years)	0.004***	0.002***	0.001**
	(0.000)	(0.000)	(0.000)
Male (yes=1)	0.107***	0.072***	0.043***
	(0.013)	(0.011)	(0.010)
Scaled labour market income		0.001*	0.005***
		(0.001)	(0.001)
Scaled non-labour market income		0.000	0.003
		(0.000)	(0.002)
Wage employment (yes=1)		0.201***	0.153***
		(0.017)	(0.015)
Casual/self-employment		0.115***	0.082***
		(0.021)	(0.019)
Number of children aged 6-16 in the household			-0.033***
			(0.005)
Number of children under 6 in the household			-0.026***
			(0.006)
Co-resident mother (yes=1)			-0.065***
			(0.009)
Co-resident father (yes=1)			-0.038**
			(0.016)
Rural location (yes=1)			-0.002
			(0.012)
(b) Conditional marginal effects			
Difference between labour market income of graduates vs. others			-0.007***
			(0.002)
Difference between non-labour income of graduates vs. others			-0.005
			(0.005)
The effect of post-secondary education at labour market income equal to:			
R1 000			0.072***
			(0.016)
R5 000			0.043***
			(0.016)
R8 000			0.019
			(0.019)
R11 000			-0.007
			(0.023)
Observations	18 913	18 913	18 913

Source: Author's own calculations using NIDS Wave 5.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. Standard errors are robust to clustering at the cluster level. Data are weighted using post-stratification weights. The dependent variable takes the value 1 if an individual remits and 0 otherwise. The income variables are scaled to represent the effect of a R1 000 change in income on the probability of remitting. Labour market income values for the conditional marginal effects were chosen based off mean values presented in Table 4.3.

The AMEs on the post-school qualification indicators and matric indicator remain significant, and the size of the AMEs for qualifications not requiring matric and post-secondary qualifications have reduced to 0.06 and 0.05 respectively. Both wage employment and casual/self-employment have a practically large and statistically significant relationship with the probability of remitting. The effect of labour market income, although statistically significant, is practically quite small: a R1 000 increase in labour market income increases the probability of remitting by 0.1 percentage points.

The results of the third regression (column 3, which introduces controls for living arrangements and includes interaction terms between the income variables and graduate status) reinforce the descriptors discussed in the previous section, namely: labour market characteristics and household structure are important determinants of who remits.[134] The statistically significant AMEs on the number of children in the household suggest a decrease in the probability of remitting of around 3 percentage points for each additional co-resident child. Similarly, having a co-resident parent decreases the probability of remitting, with AMEs of -0.07 for mothers and -0.04 for fathers.

The logic for the interpretation of interaction terms in linear models, however, does not extend to non-linear models (e.g. the probit and truncated regression models here). In non-linear models, the full interaction effect is in fact the cross-partial derivative, or cross difference, of the expected value of the dependent variable (Ai & Norton, 2003; Norton et al., 2004). Norton et al. (2004) additionally note that the z-statistic reported in non-linear regression output cannot necessarily be used to determine statistical significance of interaction terms. For these reasons, I evaluate the interaction term as the difference in *conditional* marginal effects of labour market income when graduate status is set to both 0 (other individuals) and 1 (graduates). These differences are presented in panel (b) of Table 4.7.[135] The first row of panel (b) shows that graduates' labour market income has a statistically significant weaker marginal effect (by 0.7 percentage points) on remitting than does other individuals' income. This suggests a weaker relationship between labour market income and the probability of remitting for graduates compared to other individuals. These findings suggest a reality where graduates face responsibilities to remit owing to their graduate status itself, over and above their earnings capacity.

In this third specification, the AMEs on qualifications not requiring matric and on post-secondary qualifications are further reduced in size and significance. However, given the interaction term between income and graduate status, the AME on post-secondary education in column 3 can be better understood when computed at various income levels. Table 4.7 shows

[134] Although, as alluded to earlier, there may be an element of reverse causality where graduate status affects household structure. For example, Klasen and Woolard (2008) show that the presence of a state old age pension can draw unemployed youth into a household.

[135] The results I infer about the interaction terms from this method are substantively similar to coefficients on the interaction terms in a linear probability model.

these conditional marginal effects. As income increases, the effect of post-secondary education on the probability of remitting weakens. It remains statistically significant at lower income levels, but at higher levels it is not significantly different from zero (mean income for the estimation sample is R1 768). The sign reversal around income levels of R11 000, although not statistically different from zero, is consistent with the idea that above a certain income level, individuals may have families that do not necessarily require remittance support.

It should be noted that in the second and third regressions (Table 4.7), the AME on post-secondary education is not statistically different to the coefficient on qualification not requiring matric when compared against incomplete secondary as a base category. This suggests that graduates with any type of post school qualification are more likely to remit. It is worth noting, however, that post-secondary education has a statistically significant positive probability on remitting in all specifications when compared to all other education as base category (not shown here). I now turn to explore remittance sending in more detail, specifically considering the determinants of remittance value in Table 4.8.

Table 4.8 presents the AMEs for the determinants of remittance value, conditional on remitting. The first regression in panel (a) considers the relationship between education and the value of the remittance sent, controlling only for gender and a quadratic in age. These results suggest that graduates of post-secondary education remit significantly more than those with incomplete secondary. However, once controls for labour market outcomes are introduced in columns 2 and 3 (column 3 again introduces controls for living arrangements and includes interaction terms between the income variables and graduate status), this relationship is no longer statistically significant when compared to incomplete secondary as a base category.[136]

Supporting this, panel (b) in Table 4.8 further shows that the relationship between income and remittance value (the graduate-income interaction effect) does not differ significantly by remitters' graduate status, after controlling for other remittance determinants. This would suggest that graduate remitters are not sending more relative to their 'earnings' capacity' than other remitters. However, panel (b) in Table 4.8, which also compares the MEs of post-secondary education at various labour market incomes, suggests the statistical insignificance of graduate status holds only around the mean labour market income value for graduate remitters (R11 000). Graduates earning less than this do appear to remit significantly more than other remitters earning similar levels of income.

The statistically significant negative AME on qualification not requiring matric in column 3 of Table 4.8 is also of interest. It suggests that controlling for income, individuals with these qualifications are sending less than an individual earning a similar amount with an incomplete secondary education. What is also noticeable is that the AME on matric is also negative once labour market income is controlled for, and the AME on primary schooling is positive (though

[136]It is, however significant compared to no post-secondary qualification as a base category (not shown here).

they are not significant). This may be related to the share of income going towards remittances being much higher in the groups with negative effects (Table 4.5).

Some of the other AMEs, although not directly related to my research questions, are interesting for understanding the determinants of remittance values. Posel (2001) finds that remitters in regular (wage) employment send less on average given other characteristics, and she attributes this to self-interested motives of remitting – those with more job security in their place of work are less likely to need to ensure they access the resources of the household of origin in the future. My results, however, suggest the opposite is the case. Wage employment is associated with an increase in the amount sent, although not statistically significantly so. This may stem from the fact that 73% of remitters are in wage employment and of those in any employment, 89% have wage employment.

Lastly, not only do the number of children work to lower the probability of selection into remitting, but children of all ages also reduce the value of remittances sent for those who do remit. This contrasts Posel (2001), who found that it was only the presence of children of school going age (6-18) who reduced the value of remittances sent.

Table 4.8 Truncated regression average and conditional marginal effects for the amount remitted, conditional on remitting

(a) Average marginal effects	(1)	(2)	(3)
Education (base: incomplete secondary)			
No schooling	-268.796	-232.936	-192.957
	(211.755)	(220.582)	(251.303)
Primary school	282.778	374.915	407.008
	(283.864)	(282.522)	(306.184)
Matric	36.574	-90.761	-142.941
	(190.654)	(179.635)	(222.850)
Qualification not requiring matric	-284.284*	-279.991*	-252.603*
	(158.400)	(151.612)	(135.356)
Post-secondary	641.148***	298.691	314.486
	(219.318)	(196.929)	(203.110)
Age (years)	35.319***	32.207***	34.582***
	(6.303)	(7.046)	(7.412)
Male (yes=1)	474.320***	340.029**	217.876
	(124.730)	(134.838)	(146.146)
Scaled labour market income		22.120***	25.079***
		(5.276)	(7.830)
Scaled non-labour market income		9.525*	8.080
		(5.318)	(14.516)
Wage employment (yes=1)		365.340*	234.080
		(197.277)	(202.105)
Casual/self-employment (yes=1)		230.844	120.660
		(282.482)	(270.664)
Number of children aged 6-16 in the household			-256.050**
			(105.153)
Number of children under 6 in the household			-270.548**
			(117.582)
Co-resident mother (yes=1)			-74.162
			(236.808)
Co-resident father (yes=1)			-173.588
			(260.504)
Rural location (yes=1)			-182.780
			(153.948)
(b) Conditional marginal effects			
Difference between labour market income of graduates vs. others			-0.690
			(14.375)
Difference between non-labour income of graduates vs. others			-6.354
			(41.280)
The effect of post-secondary education at labour market income equal to:			
R1 000			324.809*
			(173.630)
R5 000			328.213*
			(177.527)
R8 000			329.514*
			(186.755)
R11 000			329.430
			(203.565)
Observations	1 664	1 664	1 664

Source: Author's own calculations using NIDS Wave 5.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. Standard errors are robust to clustering at the cluster level. Data are weighted using post-stratification weights. The AMEs are calculated from coefficients of a truncated regression where the dependent variable is the positive value of cash and in-kind remittances sent in the last month. Labour market income values for the conditional marginal effects were chosen based off mean values presented in Table 4.3.

4.7. Concluding remarks

The discussion on black tax and informal, private networks of support is becoming more prominent as an increasing share of young, Black individuals access a post-secondary education and benefit from subsequently higher labour market returns. A key component of this study was thus to re-visit remittance behaviour in light of the potentially increased responsibilities graduates may face to offer financial assistance to family in need. This study, to the best of my knowledge, is also the first of its kind to use a nationally representative data set to quantitatively analyse private transfers within the frameworks of black tax and the responsibilities associated with the completion of post-secondary education in South Africa.

I set up an income-pooling test scenario to evaluate whether a unitary model of household decision making underpins remittance behaviour, but do not find evidence in this regard. The fact that remitters, whether graduates or not, are more likely to be employed and earn incomes that comprise a greater share of household income, compared to those who do not remit, may be indicative of increased bargaining power to select into remitting, or simply that, unsurprisingly, selecting into remitting is strongly determined by the availability of income to remit. However, results suggest that graduates are more likely to remit to their parents and support more people, versus other remitters in general.

Finding that the probability of remitting is higher for graduates, even once individual income is controlled for, is my first indication that graduates face a differential responsibility to remit. Second, the fact that graduates' labour market income has a weaker effect on the probability of remitting, compared to other individuals, further suggests that there is an element of responsibility that arises from graduate status alone. It appears, though, that after a certain income threshold, graduate status no longer significantly affects the probability of remittance sending. Of those who do remit, graduate remitters' income is significantly higher than that of other remitters. However, after controlling for determinants of remittance value including individual income, and conditional on remitting, graduate status is not statistically related to the amount sent.

Compared to other developing nations, South Africa has relatively low level of remittance sending. This is perhaps unsurprising given the existence of extensive social support programs. The discussion on black tax, however, does not distinguish support extended between households from support offered between individuals residing in a common household of residence. In this regard, observing that a relatively low share of households report sending remittances should not be read as empirical evidence against the prevalence of black tax or private networks of support. Where intra-household transfers are more prominent than inter-household transfers, survey data may be unable to directly elicit the true extent of private networks of support.

That said, the discussion shows the importance of considering the role of black tax in young people's lives, and how post-secondary education can affect aspirations of collective social mobility, but also individuals' economic agency. Black tax therefore has the potential

to maintain power imbalances in an unequal society, and highlights insufficient attention paid to transformative social protection.[137]

More pointedly, these results provide a descriptive overview of graduate responsibilities evidenced in remitting behaviour, and I do not attempt to identify causal effects of being a graduate on family/extended family responsibilities and black tax more generally. Nonetheless, results still have relevance for policy, including policies aiming to disrupt intergenerational social inequality, policies aimed at altering saving behaviours, and those on the funding of post-secondary education. In this latter regard, findings suggest that post-secondary education financed through a graduate tax, for example, could be seen to essentially tax Black graduates twice through increased likelihood of remittances sent. Future research looking at the effect of private transfers on intergenerational saving and consumption patterns would be a valuable addition to scholarship in this field.

[137]See also Oppel (2022) and Webb (2021).

Chapter 5
Conclusion: Main findings and reflections

The challenge of overcoming inequality in South Africa is complicated by the reality that inequality pervades even the very systems and institutions – such as post-schooling – that aspire to dismantle it. To echo what was said in Chapter 1, understanding the challenges of inequality and how they are interlinked should be a first step to informing evidence-based solutions. Post-school education can either contribute to social inequalities or be part of a remedy for reform. Whether is it one or the other is not necessarily unambiguous. As the previous chapter recognises, graduate responsibilities can contribute important social externalities, while at the same time replicating intergenerational social inequalities. In Chapter 1, I presented a cycle that visually portrayed how inequalities in post-school education can perpetuate broader socioeconomic inequalities. In order to promote a South African post-schooling system that is responsive to the needs of South African society, cycles of inequality need to be disrupted.

In this thesis, I set out to better understand three dimensions of inequality in post-schooling. By deepening an understanding of inequalities in access, achievement, and graduate realities, attention can be drawn to where there is scope to transform cycles of inequality to ones of equal opportunity.

5.1. Main findings

The first substantive contribution of this thesis involved research intended to inform the development of a sustainable, comprehensive, and progressive financial aid scheme to fund students enrolled in post-school education. In this regard, Chapter 2 proposed an empirical and conceptual methodology for analysing the NSFAS-eligible and missing middle groups based on income mobility and the notion of vulnerability.

A key value in my approach lay in operationalising the concepts of mobility, vulnerability, and economic stability to differentiate the socioeconomic circumstances of households within the context of the current post-school funding policy. My results showcased that aggregating household circumstances within the missing middle would overstate the long-run well-being of vulnerable youth and understate the relative economic security of youth I classify as relatively more stable.

While income thresholds are always likely to exist for operational purposes in social support environments, the findings of this chapter have two important considerations for reflecting on the design of support: 1) the current NSFAS funding threshold is well targeted at the those living in the most resource-constrained households, and 2) when considering extending support to the missing middle, a differentiated funding instrument could work. For example, the vulnerable missing middle looks far more likely to be engaging in formalised credit markets. This does

not dispute the fact that vulnerability to income fluctuations can be an important constraint to post-school access – even when income is above the threshold.

I turn next (in Chapter 3) to the effects of COVID-19-related closures on undergraduate students' academic performance at the University of Cape Town. Three main concerns were raised that motivated the importance of understanding potential changes in students' academic trajectories. The first was that spuriously improved or even constant (measured) academic performance during the pandemic could unintentionally impact academic performance in the long term if students proceed to higher levels without sufficient baseline knowledge and competencies. Second, and relatedly, were concerns about the quality of grades and graduation as labour market signals of ability and knowledge acquired. The third concern highlighted in this chapter was the potential for achievement gaps between students funded by NSFAS and those not funded by NSFAS to widen, owing to the fact that household inequalities were likely to impact more directly on students' learning and performance.

I find significant improvements in student performance in 2020, although the size of the improvement is practically small. While students are unlikely to be progressing more quickly through their qualifications as a result of these improvements, results nonetheless suggest that academic performance, as a usual signal of learning acquired, may be less informative. The decline observed in 2021 relative to pre-pandemic performance outcomes, together with the overall weaker correlation between performance in 2020 and performance in 2021, suggests that the improvements of 2020 were unlikely to represent true learning gains. Thus, while graduates entering the labour market from COVID-cohorts may appear to be on par academically with graduates from previous cohorts, the quality of grades and graduation as labour market signals of ability and knowledge acquired are likely to be weaker.

For institutions grappling with student success, recognising where achievement gaps between students have grown is important. Results suggest that students who were academically strong, or living in well-resourced households, were not as greatly impacted by COVID-19 closures as those students who were performing at the bottom end of the GPA distribution prior to the pandemic, or who came from lower-income socioeconomic backgrounds (as proxied by NSFAS receipt). In 2021, this gap widened further. Part of this owes to the fact that dropout between 2020 and 2021 was considerably lower than pre-pandemic times, but this heterogeneity in the extent of learning progress (or rather lack thereof) is concerning. Continued measurement of progress will be fundamental.

The discussion on black tax and informal, private networks of support is becoming more prominent as an increasing share of young African individuals access a post-secondary education and benefit from subsequently higher labour market returns. A key component of my third substantive contribution (Chapter 4) was thus to re-visit remittance behaviour in light of the potentially increased responsibilities graduates may face to offer financial assistance to family in need. This study, to the best of my knowledge, is also the first of its kind to use a nationally

representative data set to quantitatively analyse private transfers within the frameworks of black tax and the responsibilities and social externalities associated with the completion of post-secondary education in South Africa.

Unsurprisingly, selecting into remitting is strongly determined by the availability of income to remit. However, finding that the probability of remitting is higher for graduates, even once individual income is controlled for, is my first indication that graduates face a differential responsibility to remit. Second, the fact that graduates' labour market income has a weaker effect on the probability of remitting, compared to other individuals, further suggests that there is an element of responsibility that arises from graduate status alone.

These results are important for informing policy, including policies aiming to disrupt intergenerational social inequality and those on the funding of post-secondary education. In this latter regard, my results suggest that post-secondary education financed through a graduate tax, for example, could be seen to essentially tax African graduates twice through increased likelihood of remittances sent.

5.2. Final reflections

An over-arching impression from the findings above is the interconnected nature of post-schooling-related inequalities both within the system itself and society more broadly. Additionally, funding-related themes are apparent across all three chapters, whether it be issues of funding access, how funding is financed and its implications for graduate realities, or how household inequalities can impact student success (as proxied by funding receipt).

In the latter regard, relevant for meeting skills demand is that students successfully participate in their education and graduate with meaningful knowledge. The ramifications of COVID-19 on academic performance, however, could affect graduate labour market outcomes. This could be especially prejudicial for Black graduates who face responsibilities regardless of their labour market outcomes. This, in turn, is a particularly relevant consideration as government grapples with sustainably financing and expanding financial aid.

By acknowledging the existence and implications of graduate responsibilities, policy decisions can be avoided that may perpetuate inequality further, such as a graduate tax. More generally, however, graduate responsibilities call attention to the fact that where policy recognises that donations toward the public good reduce taxable income, no such recognition is given to familial support. It is thus evident that inequalities as they relate to post-schooling have layered complexities, rendering the design of suitable policy a challenging task. In building a deeper understanding of these inequalities and how they interact, however, it is hoped that empowered, informed, and complementary policy decisions can be made. Critically, a commitment to on-going research of inequalities in post-schooling will be important going forward, especially as new reforms and policy changes are implemented.

References

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, *80*(1), 123–129. https://doi.org/10.1016/S0165-1765(03)00032-6

Allais, S., Shalem, Y., & Cooper, A. (2021). Is education the answer to inequality in SA? *New Frame*. Retrieved July 15, 2022, from https://www.newframe.com/is-education-the-answer-to-inequality-in-sa/

Allington, R. L., McGill-Franzen, A., Camilli, G., Williams, L., Graff, J., Zeig, J., Zmach, C., & Nowak, R. (2010). Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology*, *31*(5), 411–427. https://doi.org/10.1080/02702711.2010.505165

Ambler, K. (2016). Bargaining with Grandma: The impact of the South African pension on household decision-making. *The Journal of Human Resources*, *51*(4), 900–932. https://doi.org/10.3368/jhr.51.4.0314-6265R1

Angrist, N., Bergman, P., Brewster, C., & Matsheng, M. (2020). *Stemming learning loss during the pandemic: A rapid randomized trial of a low-tech intervention in Botswana* (CSAE Working PaperWPS/202013). Centre for the Study of African Economies, University of Oxford. Retrieved November 19, 2021, from https://www.ssrn.com/abstract=3663098

Angrist, N., de Barros, A., Bhula, R., Chakera, S., Cummiskey, C., DeStefano, J., Floretta, J., Kaffenberger, M., Piper, B., & Stern, J. (2021). Building back better to avert a learning catastrophe: Estimating learning loss from COVID-19 school shutdowns in Africa and facilitating short-term and long-term learning recovery. *International Journal of Educational Development*, *84*. https://doi.org/10.1016/j.ijedudev.2021.102397

Ardington, C. (2021). *COVID-19 learning losses: Early grade reading in South Africa* (tech. rep.). Southern Africa Labour and Development Research Unit, University of Cape Town. Retrieved November 18, 2021, from https://fundawande.org/img/cms/news/Ardington%202021%20-%20Funda%20Wande%20EC%20learning%20losses%20report%20(24%20May%202021)_1.pdf

Ardington, C., Wills, G., & Kotze, J. (2021). COVID-19 learning losses: Early grade reading in South Africa. *International Journal of Educational Development*, *86*. https://doi.org/10.1016/j.ijedudev.2021.102480

Aristei, D. (2013). A blinder–oaxaca decomposition for double-hurdle models with an application to migrants' remittance behaviour. *Applied Economics Letters*, *20*(18), 1665–1672. https://doi.org/10.1080/13504851.2013.835468

Aucejo, E. M., French, J., Araya, M. P. U., & Zafar, B. (2020). The impact of COVID-19 on student experiences and expectations: Evidence from a survey. *Journal of Public Economics*, *191*, 104271. https://doi.org/10.1016/j.jpubeco.2020.104271

Augusta, C., & Henderson, R. D. E. (2021). Student academic integrity in online learning in higher education in the era of COVID-19. In C. Cheong, J. Coldwell-Neilson, K. MacCallum, T. Luo, & A. Scime (Eds.), *COVID-19 and Education: Learning and Teaching in a Pandemic-Constrained Environment*. Santa Rosa, California: Informing Science Press.

Azevedo, J. P., Hasan, A., Goldemberg, D., Iqbal, S. A., & Geven, K. (2020). Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates. *The World Bank Research Observer*, *36*, 1–40. https://doi.org/10.1596/1813-9450-9284

Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, *75*(299), 493–517. https://doi.org/10.2307/2228949

Becker, G. S. (1974). A theory of social interactions. *Journal of Political Economy*, *82*(6), 1063–1093. https://doi.org/https://doi.org/10.1086/260265

Bertrand, M., Mullainathan, S., & Miller, D. (2003). Public policy and extended families: Evidence from pensions in South Africa. *The World Bank Economic Review*, *17*(1), 27–50. https://doi.org/10.1093/wber/lhg014

Bettin, G., Lucchetti, R., & Zazzaro, A. (2012). Endogeneity and sample selection in a model for remittances. *Journal of Development Economics*, *99*(2), 370–384. https://doi.org/10.1016/j.jdeveco.2012.05.001

Birdsall, N. (2007). Do no harm: Aid, weak institutions and the missing middle in Africa. *Development Policy Review*, *25*(5), 575–598. https://doi.org/10.1111/j.1467-7679.2007.00386.x

Blundell, R., & Meghir, C. (1987). Bivariate alternatives to the Tobit model. *Journal of Econometrics*, *34*(1), 179–200. https://doi.org/10.1016/0304-4076(87)90072-8

Bollard, A., McKenzie, D., Morten, M., & Rapoport, H. (2011). Remittances and the brain drain revisited: The microdata show that more educated migrants remit more. *The World Bank Economic Review*, *25*(1), 132–156. https://doi.org/10.1093/wber/lhr013

Bourguignon, F., Browning, M., & Chiappori, P.-A. (2009). Efficient intra-household allocations and distribution factors: Implications and identification. *Review of Economic Studies*, *76*(2), 503–528. https://doi.org/10.1111/j.1467-937X.2008.00525.x

Bourguignon, F., Browning, M., Chiappori, P.-A., & Lechene, V. (1993). Intra household allocation of consumption: A model and some evidence from French data. *Annales d'Économie et de Statistique*, *29*, 137–156.

Bourguignon, F., & Chiappori, P.-A. (1992). Collective models of household behavior: An introduction. *European Economic Review*, *36*(2-3), 355–364. https://doi.org/https://doi.org/10.1016/0014-2921(92)90091-A

Branson, N., Ardington, C., Lam, D., & Leibbrandt, M. (2013). *Changes in education, employment and earnings in South Africa – A cohort analysis* (SALDRU Working Paper No. 105). Southern Africa Labour and Development Research Unit, University of Cape Town.

Branson, N., Culligan, S., & Ingle, K. (2020). *Developing Siyaphambili: A Stronger South African Nation Website. Moving towards a unified goal to combat inequality and unemployment* (SALDRU Working Paper No. 20/01). Southern Africa Labour and Development Research Unit, University of Cape Town.

Branson, N., & Kahn, A. (2019). The post matriculation enrolment decision: Do public TVET colleges provide students with a viable alternative? In M. Rogan (Ed.), *Post-school education and the labour market in South Africa* (pp. 18–37). HSRC Press.

Branson, N., & Wittenberg, M. (2019). *Longitudinal and cross-sectional weights in the NIDS data 1-5* (NIDS Technical Paper No. 9). Southern Africa Labour and Development Research Unit, University of Cape Town. Cape Town.

Browning, M., Bourguignon, F., Chiappori, P.-A., & Lechene, V. (1994). Income and outcomes: A structural model of intrahousehold allocation. *Journal of Political Economy*, *102*(6), 1067–1096. https://doi.org/10.1086/261964

Burger, R., & Kreuser, F. (2017). *Unequal partners: The determinants and consequences of latent female household bargaining power* (REDI3x3 Working paper No. 44).

Caldwell, J. C. (1965). Extended family obligations and education: A study of an aspect of demographic transition amongst Ghanaian university students. *Population Studies*, *19*(2), 183–199. https://doi.org/10.1080/00324728.1965.10406012

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Cappellari, L., & Jenkins, S. P. (2003). Multivariate probit regression using simulated maximum likelihood. *The Stata Journal*, *3*(3), 278–294. https://doi.org/10.1177/1536867X0300300305

Cappellari, L., & Jenkins, S. P. (2004). Modelling low income transitions. *Journal of Applied Econometrics*, *19*(5), 593–610. https://doi.org/10.1002/jae.778

Carrell, S. E., Malmstrom, F. V., & West, J. E. (2008). Peer effects in academic cheating. *The Journal of Human Resources*, *43*(1), 173–207. https://doi.org/10.3368/jhr.43.1.173

Carter, M. R., & May, J. (2001). One kind of freedom: Poverty dynamics in post-apartheid South Africa. *World Development*, *29*(12), 1987–2006. https://doi.org/10.1016/S0305-750X(01)00089-4

Centre for Teaching and Learning (UFS). (2020). UCT SAULM Data Summary.

Chaka, C. (2020). Higher education institutions and the use of online instruction and online tools and resources during the COVID-19 outbreak - An online review of selected U.S. and SA's universities. *Preprint*. https://doi.org/10.21203/rs.3.rs-61482/v1

Chatterjee, A., Czajka, L., & Gethin, A. (2022). Wealth inequality in South Africa, 1993–2017. *The World Bank Economic Review*, *36*(1), 19–36. https://doi.org/10.1093/wber/lhab012

Chauke, C. (2019). Taking the sting out of black tax. In N. Mhlongo (Ed.), *Black Tax* (pp. 196–201). Johannesburg & Cape Town: Jonathan Ball Publishers.

Chiappori, P.-A. (1992). Collective labor supply and welfare. *Journal of Political Economy*, *100*(3), 437–467. https://doi.org/10.1086/261825

Chiteji, N. S., & Hamilton, D. (2002). Family connections and the black-white wealth gap among middle-class families. *The Review of Black Political Economy*, *30*(1), 9–28. https://doi.org/10.1007/BF02808169

Commission of Inquiry into Higher Education and Training. (2017). *Report of the Commission of Enquiry into Higher Education and Training to the President of the Republic of South Africa* (tech. rep.). Retrieved June 25, 2020, from https://www.gov.za/sites/default/files/gcis_document/201711/commissioninquiryhighereducationreport.pdf

Council on Higher Education (South Africa). (2013). *The Higher Education Qualifications Sub-Framework* [OCLC: 881681732]. Council on Higher Education (CHE). Retrieved August 13, 2022, from https://nr-online-1.che.ac.za/html_documents/6.PUB_HEQSF.PDF

Cox, D., & Fafchamps, M. (2008). Extended family and kinship networks: Economic insights and evolutionary directions. In T. P. Schultz & J. A. Strauss (Eds.), *Handbook of Development Economics* (pp. 3711–3784).

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, *39*(5), 829–844. https://doi.org/10.2307/1909582

Culligan, S. (2022). *Using census, institutional and geospatial data to estimate the socio-economic profile of post-school students by institutional type* (Master's thesis). University of Cape Town. South Africa.

Czerniewicz, L. (2020). University shutdowns – What we learnt from 'going online'. *University World News*. Retrieved November 22, 2021, from https://www.universityworldnews.com/post.php?story=20200325160338881

Dandara, C., Chimusa, E. R., & Wonkam, A. (2017). South African student protests, 2015-2016: The aftermath viewed through Medical Science Honours students at the University of Cape Town. *South African Medical Journal*, *107*(9), 723–723. https://doi.org/http://dx.doi.org/10.7196/samj.2017.v107i9.12714

Daya, R., & Mpete, L. (2017). How to escape the 'black tax'. *The Citizen*. Retrieved January 13, 2020, from https://www.pressreader.com/south-africa/the-citizen-gauteng/20170213/282011852109424

Department of Higher Education and Training. (2013). White paper for post-school education and training. Building an expanded, effective and integrated post-school system. Retrieved July 15, 2022, from https://www.dhet.gov.za/SiteAssets/Latest%20News/White%20paper%20for%20post-school%20education%20and%20training.pdf

Department of Higher Education and Training. (2020). *Revised Strategic Plan 2020-2025* (tech. rep.). Retrieved March 15, 2022, from https://www.dhet.gov.za/SiteAssets/DHET%20Strategic%20Plan%202020.pdf

Department of Higher Education and Training. (2021). Post-School Education and Training Monitor: Macro-Indicator Trends. Retrieved September 29, 2022, from https://www.dhet.gov.za/Planning%20Monitoring%20and%20Evaluation%20Coordination/Post-School%20Education%20and%20Training%20Monitor%20-%20Macro-Indicator%20Trends%20-%20March%202021.pdf

Department of Higher Education and Training. (2022). Statistics on Post-School Education and Training in South Africa 2020. Retrieved September 29, 2022, from https://www.dhet.gov.za/DHET%20Statistics%20Publication/Statistics%20on%20Post-School%20Education%20and%20Training%20in%20South%20Africa%202020.pdf

Devadoss, S., & Foltz, J. (1996). Evaluation of factors influencing student class attendance and performance. *American Journal of Agricultural Economics*, *78*(3), 499–507. https://doi.org/https://doi.org/10.1016/j.chieco.2021.101629

di Falco, S., & Bulte, E. (2011). A dark side of social capital? Kinship, consumption, and savings. *Journal of Development Studies*, *47*(8), 1128–1151. https://doi.org/10.1080/00220388.2010.514328

Duflo, E. (2003). Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in South Africa. *The World Bank Economic Review*, *17*(1), 1–25. https://doi.org/10.1093/wber/lhg013

Duhaut, A., Kondylis, F., Loeser, J., & Piza, C. (2021a). DiD you see Beta? Beta who? Part 1. Retrieved June 30, 2022, from https://blogs.worldbank.org/impactevaluations/did-you-see-beta-beta-who-part-1

Duhaut, A., Kondylis, F., Loeser, J., & Piza, C. (2021b). DiD you see Beta? Beta who? Part 2. Retrieved June 30, 2022, from https://blogs.worldbank.org/impactevaluations/did-you-see-beta-beta-who-part-2

du Toit, A., & Neves, D. (2006). *Vulnerability and social protection at the margins of the formal economy. Case studies from Khayelitsha and the Eastern Cape*. United States Agency for International Development.

Faculty of Commerce. (2022). Faculty of Commerce Undergraduate Studies. Retrieved August 4, 2022, from http://www.students.uct.ac.za/sites/default/files/image_tool/images/434/study/handbooks/2022/2022_COM_UG_Handbook.pdf

Finn, A., & Leibbrandt, M. (2017). *The dynamics of poverty in South Africa* (SALDRU Working Paper No. 174/NIDS Discussion Paper No. 2016/1). Southern Africa Labour and Development Research Unit, University of Cape Town.

Fongwa, S. N. (2019). Interrogating the public good versus private good dichotomy: 'black tax' as a higher education public good. *Compare: A Journal of Comparative and International Education*, 1–16. https://doi.org/10.1080/03057925.2019.1651194

Frankenberg, E., Lillard, L., & Willis, R. J. (2002). Patterns of intergenerational transfers in Southeast Asia. *Journal of Marriage and Family*, *64*(3), 627–641. https://doi.org/10.1111/j.1741-3737.2002.00627.x

Garrod, N., & Wildschut, A. (2021). How large is the missing middle and what would it cost to fund? *Development Southern Africa*, *38*(3), 484–491. https://doi.org/10.1080/0376835X.2020.1796594

Gonzalez, T., de la Rubia, M., Hincz, K., Comas-Lopez, M., Subirats, L., Fort, S., & Sacha, G. (2020). Influence of COVID-19 confinement on students' performance in higher education (H. Xie, Ed.). *PLoS ONE*, *15*(10), e0239490. https://doi.org/10.1371/journal.pone.0239490

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

Gourieroux, C., & Monfort, A. (1996). *Simulation-based econometric methods*. Oxford university press.

Gumede, W. (2022). The problem of funding SA students can be solved. Here's how. *Sunday Times*. Retrieved August 1, 2022, from https://www.timeslive.co.za/sunday-times-daily/opinion-and-analysis/2022-05-09-william-gumede--the-problem-of-funding-sa-students-can-be-solved-heres-how/

Gummerson, E., & Schneider, D. (2013). Eat, drink, man, woman: Gender, income share and household expenditure in South Africa. *Social Forces*, *91*(3), 813–836. https://doi.org/10.1093/sf/sos173

Gustafsson, M., & Nuga Deliwe, C. (2020). *How is the COVID-19 pandemic affecting educational quality in South Africa?* (NIDS-CRAM Insight Brief). Department of Economics, University of Stellenbosch.

Haddad, L. J., & Kanbur, S. M. R. (1990). *Are better-off households more unequal or less unequal?* World Bank Publications.

Hall, K. (2016). *Maternal and child migration in post-apartheid South Africa: Evidence from the NIDS panel study* (SALDRU Working Paper No. 178/NIDS Discussion Paper No. 2016/5). Southern Africa Labour and Development Research Unit, University of Cape Town.

Harper, S., & Seekings, J. (2010). *Claims on and obligations to kin in Cape Town, South Africa* (CSSR Working Paper No. 272). Centre for Social Science Research, University of Cape Town. Cape Town.

Heckman, J. J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time - discrete data stochastic process. In C. Manski & D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications* (pp. 179–195). MIT Press.

Hutton, P. A. (2006). Understanding student cheating and what educators can do about it. *College Teaching*, *54*(1), 171–176. https://doi.org/10.3200/CTCH.54.1.171-176

Iglesias-Pradas, S., Hernández-García, Á., Chaparro-Peláez, J., & Prieto, J. L. (2021). Emergency remote teaching and students' academic performance in higher education during the COVID-19 pandemic: A case study. *Computers in Human Behavior*, *119*, 106713. https://doi.org/10.1016/j.chb.2021.106713

Jacobs, L., Moolman, A., & de Beer, E. (2019). #Feesmustfall and beyond: Towards a sustainable national student loan regulatory framework. *South African Journal of Higher Education*, *33*(1), 127–143. https://doi.org/10.20853/33-1-1336

Jenkins, S. P. (2011). Modelling low income transitions: Estimates from a Markov model. In *Changing Fortunes: Income Mobility and Poverty Dynamics in Britain* (pp. 330–356). OUP Oxford.

Jensen, R. T. (2003). Do private transfers 'displace' the benefits of public transfers? Evidence from South Africa. *Journal of Public Economics*, *88*(1), 89–112. https://doi.org/10.1016/S0047-2727(02)00085-3

Jones, A. M. (1989). A double-hurdle model of cigarette consumption. *Journal of Applied Econometrics*, *4*(1), 23–39. https://doi.org/10.1002/jae.3950040103

Keswell, M., & Poswell, L. (2004). Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence. *South African Journal of Economics*, *72*(4), 834–860. https://doi.org/10.1111/j.1813-6982.2004.tb00136.x

Klasen, S., & Woolard, I. (2008). Surviving unemployment without state support: Unemployment and household formation in South Africa. *Journal of African Economies*, *18*(1), 1–51. https://doi.org/10.1093/jae/ejn007

LaFave, D., & Thomas, D. (2017). Extended families and child well-being. *Journal of Development Economics*, *126*, 52–65. https://doi.org/10.1016/j.jdeveco.2016.11.006

Lam, D., Ardington, C., Branson, N., & Leibbrandt, M. (2013). *Credit constraints and the racial gap in post-secondary education in South Africa* (NBER Working Paper Series No. 19607). National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w19607

Lange, L. (2021). Framework for Undergraduate Teaching and Learning in 2022. Retrieved July 28, 2022, from https://www.news.uct.ac.za/downloads/comms/2022/SENATE-Framework-

for-Teaching.pdf?utm_source=mailer&utm_medium=email&utm_campaign=DVC%3A+Teaching+%26+Learning&utm_term=https%3A%2F%2Fwww.news.uct.ac.za%2Fdownloads%2Fcomms%2F2022%2FSENATE-Framework-for-Teaching.pdf

Levin, H. M., & Tsang, M. C. (1987). The economics of student time. *Economics of Education Review*, *6*(4), 357–364. https://doi.org/10.1016/0272-7757(87)90019-7

Levy, B., Hirsch, A., & Woolard, I. (2014). *South Africa's evolving political settlement in comparative perspective* (SALDRU Working Paper No. 138). Southern Africa Labour and Development Research Unit, University of Cape Town.

Lillard, L. A., & Willis, R. J. (1997). Motives for interqenerational transfers: Evidence from Malaysia. *Demography*, *34*(1), 115–134.

López-Calva, L. F., & Ortiz-Juarez, E. (2014). A vulnerability approach to the definition of the middle class. *The Journal of Economic Inequality*, *12*(1), 23–47. https://doi.org/10.1007/s10888-012-9240-5

Lundberg, S., & Pollak, R. A. (1993). Separate spheres bargaining and the marriage market. *Journal of Political Economy*, *101*(6), 988–1010. https://doi.org/10.1086/261912

Mafolo, K., & Shoba, S. (2021). The 'Owl' is watching varsity exam cheats, but tech-savvy few don't give a hoot. *Daily Maverick*. Retrieved November 23, 2021, from https://www.dailymaverick.co.za/article/2021-11-11-the-owl-is-watching-varsity-exam-cheats-but-tech-savvy-few-dont-give-a-hoot/

Magubane, N. N. (2016). *Black tax: The emerging middle class reality* (Master of Business Administration Thesis.). Gordon Institute of Business Science, University of Pretoria.

Maitra, P., & Ray, R. (2003). The effect of transfers on household expenditure patterns and poverty in South Africa. *Journal of Development Economics*, *71*(1), 23–49. https://doi.org/10.1016/S0304-3878(02)00132-3

Maldonado, J. E., & Witte, K. D. (2020). *The effect of school closures on standardised student test outcomes* (Discussion Paper Series DPS20.17). Leuven Economics of Education Research (LEER), KU Leuven.

Mangoma, A., & Wilson-Prangley, A. (2019). Black Tax: Understanding the financial transfers of the emerging black middle class. *Development Southern Africa*, *36*(4), 443–460. https://doi.org/10.1080/0376835X.2018.1516545

Manser, M., & Brown, M. (1980). Marriage and household decision-making: A bargaining analysis. *International Economic Review*, *21*(1), 31–44. https://doi.org/10.2307/2526238

Maqetuka, M. (2019). Black millennials don't have it all wrong. In N. Mhlongo (Ed.), *Black Tax* (pp. 250–257). Johannesburg & Cape Town: Jonathan Ball Publishers.

Maqhina, M. (2020). NSFAS fails to procure promised laptops for students. *Independent Online [IOL]*. Retrieved June 30, 2021, from https://www.iol.co.za/news/politics/nsfas-fails-to-procure-promised-laptops-for-students-da0c746f-6800-4d80-ac49-3d4234512e31

Marquard, S., Sukaina, W., Soraya, L., Kende, K., & Deacon, A. (2020). UCT ERT Student Experience Survey 2020. https://open.uct.ac.za/bitstream/handle/11427/34089/UCT%20ERT%20Student%20Experience%20Survey%202020.pdf?sequence=1&isAllowed=y

Mbah, M. F. (2014). The dilemma of graduate unemployment within a context of poverty, scarcity and fragile economy: Are there lessons for the university? *International Journal of Economics and Finance*, *6*(12), 27–36. https://doi.org/10.5539/ijef.v6n12p27

Mbewe, S., & Woolard, I. (2016). *Cross-Sectional features of wealth inequality in South Africa: Evidence from The National Income Dynamics Study* (SALDRU Working Paper No. 185/NIDS Discussion Paper No. 2016/12). Southern Africa Labour and Development Research Unit, University of Cape Town.

Mhlongo, N. (Ed.). (2019). *Black Tax*. Johannesburg & Cape Town: Jonathan Ball Publishers.

Mussa, R. (2013). Rural–urban differences in parental spending on children's primary education in Malawi. *Development Southern Africa*, *30*(6), 789–811. https://doi.org/10.1080/0376835X.2013.859066

Mzobe, S. (2019). The hopes and dreams of black parents. In N. Mhlongo (Ed.), *Black Tax* (pp. 166–178). Johannesburg & Cape Town: Jonathan Ball Publishers.

National Planning Commission. (2020). A Review of the National Development Plan 2030: Advancing implementation towards a more capable nation. Retrieved September 27, 2022, from https://www.nationalplanningcommission.org.za/assets/Documents/NDP%20REVIEW.pdf

National Treasury. (2022). Budget Review 2022. Retrieved August 1, 2022, from http://www.treasury.gov.za/documents/national%20budget/2022/review/FullBR.pdf

Newman, C., Henchion, M., & Matthews, A. (2003). A double-hurdle model of Irish household expenditure on prepared meals. *Applied Economics*, *35*(9), 1053–1061. https://doi.org/10.1080/0003684032000079170

Nkosi, B. (2021). Learning to cheat (remotely) on a massive scale. *Independent Online [IOL]*. Retrieved November 23, 2021, from https://www.iol.co.za/capetimes/news/learning-to-cheat-remotely-on-a-massive-scale-9da3968f-eab7-41c3-b3d3-f166cf348e60

Norton, E. C., Wang, H., & Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *The Stata Journal: Promoting communications on statistics and Stata*, *4*(2), 154–167. https://doi.org/10.1177/1536867X0400400206

NSFAS. (2016). NSFAS Submission to the Fees Commission. https://www.nsfas.org.za/content/publications/NSFAS%20Submission%20to%20the%20Fees%20Commission%2020160724.pdf

Nyathi, S. (2019). The burden of black tax can only be alleviated by generational wealth. In N. Mhlongo (Ed.), *Black Tax* (pp. 238–249). Johannesburg & Cape Town: Jonathan Ball Publishers.

Oppel, A. (2021). Normalizing necessity? Support networks and racial inequality in Namibia. *World Development*, *147*, 1–15. https://doi.org/10.1016/j.worlddev.2021.105649

Oppel, A. (2022). Patterns and lived realities: Exploring informal social protection across race and education. *International Journal of Social Welfare*, *31*(4), 407–420. https://doi.org/10.1111/ijsw.12548

Pillay, A., Khosa, M., Sheik, A., Campbell, B., Mthembu, B., & Nyika, N. (2021). How home contexts of South African university students shape their experiences of emergency remote teaching and learning. *Student Success*, *12*(2). https://doi.org/10.5204/ssj.1779

Posel, D. (2001a). How do households work? Migration, the household and remittance behaviour in South Africa. *Social Dynamics*, *27*(1), 165–189. https://doi.org/10.1080/02533950108458709

Posel, D. (2001b). Intra-family transfers and income-pooling: A study of remittances in Kwazulu-Natal. *South African Journal of Economics*, *69*(3), 501–528. https://doi.org/10.1111/j.1813-6982.2001.tb00023.x

Posel, D. (2016). *Inter-household transfers in South Africa: Prevalence, patterns and poverty* (SALDRU Working Paper No. 180/NIDS Discussion Paper No. 2016/7). Southern Africa Labour and Development Research Unit, University of Cape Town. Cape Town.

Posel, D., Fairburn, J. A., & Lund, F. (2006). Labour migration and households: A reconsideration of the effects of the social pension on labour supply in South Africa. *Economic Modelling*, *23*(5), 836–853. https://doi.org/10.1016/j.econmod.2005.10.010

Quisumbing, A. R., & Maluccio, J. A. (2003). Resources at marriage and intrahousehold allocation: Evidence from Bangladesh, Ethiopia, Indonesia, and South Africa. *Oxford Bulletin of Economics and Statistics*, *65*(3), 283–327. https://doi.org/10.1111/1468-0084.t01-1-00052

Rahmani, I., & Wooldridge, J. (2019). Model-selection tests for complex survey samples. In K. P. Huynh, D. T. Jacho-chávez, & G. Tripathi (Eds.), *Advances in Econometrics* (pp. 109–135). Emerald Publishing Limited. https://doi.org/10.1108/S0731-905320190000039009

Rodriguez-Planas, N. (2020). *Hitting where it hurts most: Covid-19 and low-income urban college students* (IZA Discussion Papers No. 13644). Institute of Labor Economics (IZA), Bonn.

Rodriguez-Planas, N. (2021). *COVID-19 and college academic performance: A longitudinal analysis* (IZA Discussion Papers No. 14113). Institute of Labor Economics (IZA), Bonn.

Roodman, D. (2011). Fitting fully observed recursive mixed-process models with cmp. *The Stata Journal*, *11*(2), 159–206. https://doi.org/10.1177/1536867X1101100202

Sagner, A., & Mtati, R. Z. (1999). Politics of pension sharing in urban South Africa. *Ageing and Society*, *19*(4), 393–416. https://doi.org/10.1017/S0144686X99007424

SALDRU. (2017). Understanding the National Student Financial Aid Scheme: SALDRU information sheet. http://opensaldru.uct.ac.za/bitstream/handle/11090/861/NSFAS_14Feb_final.pdf?sequence=1

Salisbury, T. (2016). Education and inequality in South Africa: Returns to schooling in the post-apartheid era. *International Journal of Educational Development*, *46*, 43–52. https://doi.org/https://doi-org.ezproxy.uct.ac.za/10.1016/j.ijedudev.2015.07.004

Salmi, J. (2020). COVID's lessons for global higher education: Coping with the present while building a more equitable future.

Schirmer, S., & Nkomana, S. (2021). *South Africa's NEETs crisis: Why we are failing to connect young people to work*. (tech. rep.). The Centre for Development and Enterprise (CDE).

Schotte, S., Zizzamia, R., & Leibbrandt, M. (2018). A poverty dynamics approach to social stratification: The South African case. *World Development*, *110*, 88–103. https://doi.org/https://doi.org/10.1016/j.worlddev.2018.05.024

Schotte, S., Zizzamia, R., & Leibbrandt, M. (2022). Snakes and ladders and loaded dice: Poverty dynamics and inequality in South Africa between 2008 and 2017. *South African Journal of Economics*, *90*(2), 214–242. https://doi.org/10.1111/saje.12308

Sinning, M. G. (2011). Determinants of savings and remittances: empirical evidence from immigrants to Germany. *Review of Economics of the Household*, *9*(1), 45–67. https://doi.org/10.1007/s11150-009-9082-5

South African Government News Agency. (2022). NSFAS 2022 funding reaches 691 432 students. *SAnews.gov.za*. Retrieved August 1, 2022, from https://www.sanews.gov.za/south-africa/nsfas-2022-funding-reaches-691-432-students

Spaull, N. (2019). Equity: A price too high to pay? In *South african schooling: The enigma of inequality* (pp. 1–24). Springer.

Stats SA. (2017). Consumer Price Index: March 2017. (Statistical release P0141). Retrieved November 11, 2021, from http://www.statssa.gov.za/publications/P0141/P0141December2017.pdf

Stats SA. (2019). National Poverty Lines. (Statistical release P0310.1). http://www.statssa.gov.za/publications/P03101/P031012019.pdf

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006

The World Bank. (2018). *Overcoming poverty and inequality in South Africa. An assessment of drivers, constraints and opportunities*. Retrieved August 14, 2020, from http://documents1.worldbank.org/curated/en/530481521735906534/pdf/124521-REV-OUO-South-Africa-Poverty-and-Inequality-Assessment-Report-2018-FINAL-WEB.pdf

Thomas, D. (1990). Intra-household resource allocation: An inferential approach. *The Journal of Human Resources*, *25*(4), 635–664. https://doi.org/10.2307/145670

Thomas, D. (1996). Education across generations in South Africa. *The American Economic Review*, *86*(2), 330–334.

Tikly, L., & Barrett, A. M. (2011). Social justice, capabilities and the quality of education in low income countries. *International Journal of Educational Development*, *31*(1), 3–14. https://doi.org/10.1016/j.ijedudev.2010.06.001

Train, K. (2003). *Discrete choice methods with simulation*. Cambridge University Press.

UCT. (2020). Teaching and Learning Report 2020. Retrieved July 6, 2022, from http://www.ched.uct.ac.za/sites/default/files/image_tool/images/185/UCT_2020_T%26L-Report_Approved-by-Council_4-Dec-2021.pdf

Universities South Africa [USAf]. (2016). Universities Funding in South Africa; A Fact Sheet. Retrieved April 13, 2020, from http://www.uct.ac.za/usr/news/downloads/2016/UniversitiesFundingSouthAfrica_FactSheet.pdf

van Broekhuizen, H. (2016). *Graduate unemployment and Higher Education Institutions in South Africa* (Stellenbosch Economic Working Papers No. 08/16). Department of Economics and The Bureau for Economic Research, University of Stellenbosch.

van Schalkwyk, F. (2021). Reflections on the public university sector and the covid-19 pandemic in South Africa. *Studies in Higher Education*, *46*(1), 44–58. https://doi.org/10.1080/03075079.2020.1859682

Visagie, J., & Posel, D. (2013). A reconsideration of what and who is middle class in South Africa. *Development Southern Africa*, *30*(2), 149–167. https://doi.org/10.1080/0376835X.2013.797224

Vogel, A., & Korinek, K. (2012). Passing by the girls? Remittance allocation for educational expenditures and social inequality in Nepal's households 2003–2004. *International Migration Review*, *46*(1), 61–100. https://doi.org/10.1111/j.1747-7379.2012.00881.x

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, *57*(2), 307–333. https://doi.org/10.2307/1912557

Webb, C. (2021). Liberating the family: Debt, education and racial capitalism in South Africa. *Environment and Planning D: Society and Space*, *39*(1), 85–102. https://doi.org/10.1177/0263775820942522

Whitelaw, E., Culligan, S., & Branson, N. (2020). *Student ability to learn at home. An introductory look at student access to remote learning resources* (tech. rep.). Southern Africa Labour and Development Research Unit, University of Cape Town.

Wildschut, A., Rogan, M., & Mncwango, B. (2020). Transformation, stratification and higher education: Exploring the absorption into employment of public financial aid beneficiaries

across the South African higher education system. *Higher Education*, *79*(6), 961–979. https://doi.org/10.1007/s10734-019-00450-z

Wilson, F. (2011). Historical roots of inequality in South Africa. *Economic History of Developing Regions*, *26*(1), 1–15. https://doi.org/10.1080/20780389.2011.583026

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT press.

Zizzamia, R., Schotte, S., & Leibbrandt, M. (2019). *Snakes and ladders and loaded dice: Poverty dynamics and inequality in South Africa between 2008-2017* (SALDRU Working Paper Number No. 235/NIDS Discussion Paper No. 2019/2). Southern Africa Labour and Development Research Unit, University of Cape Town.

Zizzamia, R., Schotte, S., Leibbrandt, M., & Ranchhod, V. (2016). *Vulnerability and the middle class in South Africa* (SALDRU Working Paper No. 188/NIDS Discussion Paper No. 2016/15). Southern Africa Labour and Development Research Unit, University of Cape Town.

Data

Department of Higher Education and Training. Higher Education Management Information System, 2012 to 2020 [data set]. Pretoria: Department of Higher Education and Training [producer]. [privately distributed].

Southern Africa Labour and Development Research Unit. National Income Dynamics Study 2017, Wave 5 [dataset]. Version 1.0.0 Pretoria: Department of Planning, Monitoring, and Evaluation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2018. Cape Town: DataFirst [distributor], 2018. https://doi.org/10.25828/fw3h-v708

Southern Africa Labour and Development Research Unit. National Income Dynamics Study 2014-2015, Wave 4 [dataset]. Version 2.0.0 Pretoria: Department of Planning, Monitoring, and Evaluation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2018. Cape Town: DataFirst [distributor], 2018. https://doi.org/10.25828/f4ws-8a78

Southern Africa Labour and Development Research Unit. National Income Dynamics Study 2012, Wave 3 [dataset]. Version 3.0.0 Pretoria: Department of Planning, Monitoring, and Evaluation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2018. Cape Town: DataFirst [distributor], 2018. https://doi.org/10.25828/7pgq-q106

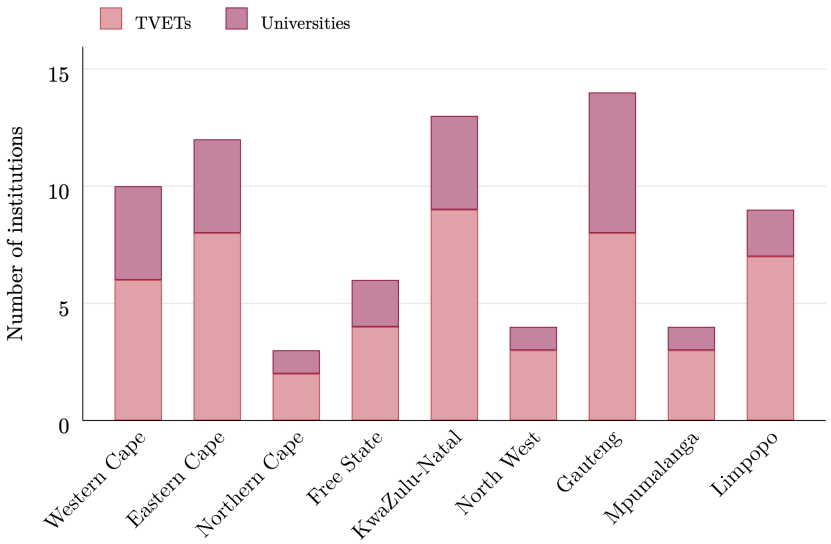
Southern Africa Labour and Development Research Unit. National Income Dynamics Study 2010-2011, Wave 2 [dataset]. Version 4.0.0 Pretoria: Department of Planning, Monitoring, and Evaluation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2018. Cape Town: DataFirst [distributor], 2018. https://doi.org/10.25828/j1h1-5m16

Southern Africa Labour and Development Research Unit. National Income Dynamics Study 2008, Wave 1 [dataset]. Version 7.0.0 Pretoria: Department of Planning, Monitoring, and Evaluation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2018. Cape Town: DataFirst [distributor], 2018. https://doi.org/10.25828/e7w9-m033

University of Cape Town [UCT]. UCT 2016-2021 [dataset]. Cape Town: Institutional Planning Department, University of Cape Town, 2022. [privately distributed].

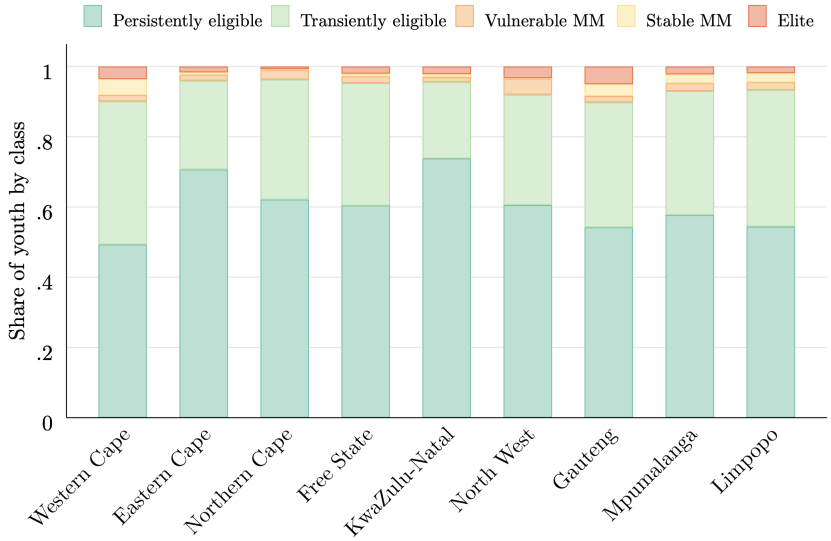
Appendix A

Figure A.1 Number of public universities and colleges by province, 2017



Source: Author's own calculations.

Figure A.2 Share of youth by funding class and province



Source: Author's own calculations using NIDS Waves 1-5 pooled sample (post-stratified weights).

Table A.1 Transitions between NSFAS classifications

Household income (time t-1)	Household income (time t)				
	Below R122 000	R122 000- R350 000	R350 000- R600 000	Above R600 000	Total
Below R122 000	86.27	12.64	0.87	0.22	100
R122 000 to R350 000	32.09	55.49	9.48	2.94	100
R350 000 to R600 000	14.51	32.06	29.63	23.80	100
Above R600 000	5.04	21.53	24.97	48.46	100
Total	69.38	22.55	4.67	3.40	100

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratified weights from the base period and corrected for panel attrition.

Table A.2 Multivariate probit: all equations (coefficient estimates)

	Transition eqn.		Initial	Retention
Base period covariate	Remain	Fall	conditions	eqn.
Characteristics of the household head				
Race (base: White)				
African	0.444**	0.702***	0.847***	0.557***
	(0.183)	(0.207)	(0.103)	(0.076)
Coloured	0.315*	0.451**	0.457***	0.446***
	(0.182)	(0.213)	(0.132)	(0.083)
Asian/Indian	0.548**	0.054	0.117	0.068
	(0.252)	(0.284)	(0.207)	(0.147)
Age of household head	-0.001	0.013**	-0.004	0.003**
	(0.003)	(0.006)	(0.003)	(0.001)
Household head is female (yes=1)	0.105*	0.245*	0.229***	0.011
	(0.061)	(0.135)	(0.060)	(0.027)
Household head is employed (yes=1)	0.123	0.218	0.008	0.048
	(0.088)	(0.155)	(0.077)	(0.037)
Highest education (base: Incomplete secondary)				
Missing	-0.537*	-2.488***	0.155	0.095
	(0.291)	(0.488)	(0.266)	(0.186)
No schooling	0.263*	-0.938**	0.504***	-0.128***
	(0.135)	(0.462)	(0.140)	(0.042)
Primary	0.052	-0.177	0.324***	-0.030
	(0.098)	(0.402)	(0.114)	(0.034)
Matric	-0.332***	-0.056	-0.184*	-0.183***
	(0.101)	(0.210)	(0.104)	(0.047)
Post-school qualification	-0.506***	-0.286	-0.434***	-0.185***
	(0.101)	(0.181)	(0.081)	(0.044)
Household characteristics				
Household employment rate	-0.379***	-0.367	-0.690***	-0.041
	(0.110)	(0.232)	(0.105)	(0.052)
At least one resident has financial assets (yes=1)	-0.038	-0.237	0.101	0.088***
	(0.092)	(0.266)	(0.087)	(0.028)
At least one resident has pension assets (yes=1)	-0.381***	-0.060	-0.405***	-0.109*
	(0.098)	(0.137)	(0.070)	(0.061)
Household member owns dwelling (yes=1)	-0.010	0.088	-0.539***	0.202***
	(0.079)	(0.207)	(0.088)	(0.034)
Household has livestock assets (yes=1)	0.288**	1.053***	0.095	0.096***
	(0.125)	(0.356)	(0.152)	(0.037)
Share of durables owned	-1.162***	1.358**	-2.559***	-0.010
	(0.377)	(0.551)	(0.266)	(0.128)
Number of rooms in house	-0.038***	-0.043*	-0.072***	0.002
	(0.014)	(0.024)	(0.013)	(0.007)
Household has access to electricity (yes=1)	-0.077	-0.274	0.175*	0.025
	(0.097)	(0.285)	(0.098)	(0.033)
Piped water on site (yes=1)	0.004	0.094	0.194*	0.068**
	(0.105)	(0.377)	(0.105)	(0.031)
Household has a flush toilet (yes=1)	-0.220*	-0.161	-0.341***	-0.033
	(0.133)	(0.243)	(0.115)	(0.036)
Number of household residents	-0.064**	0.024	-0.087***	-0.015
	(0.027)	(0.055)	(0.024)	(0.011)
Household dependency ratio	0.046	-0.061	0.170***	0.046**
	(0.057)	(0.147)	(0.061)	(0.019)

Continuation of Table A.2

	Transition eqn.		Initial	Retention
Base period covariate	Remain	Fall	conditions	eqn.
Number of children under 6	0.095	0.230*	0.074	0.044*
	(0.060)	(0.127)	(0.050)	(0.022)
Number of children aged 6-18	0.011	0.011	0.044	0.068***
	(0.036)	(0.093)	(0.043)	(0.017)
Location (base: Rural)				
Urban	-0.128	-0.026	-0.160	-0.101***
	(0.147)	(0.181)	(0.121)	(0.038)
Farm	0.035	-0.177	0.360	-0.093
	(0.175)	(0.370)	(0.255)	(0.061)
Year in time t (base: 2010)				
2012	0.169	0.054	-0.180**	0.444***
	(0.111)	(0.170)	(0.082)	(0.039)
2014	0.135	-0.338*	0.025	0.545***
	(0.090)	(0.187)	(0.083)	(0.037)
2017	0.252***	-0.229	0.078	0.470***
	(0.089)	(0.164)	(0.082)	(0.040)
Province (base: Western Cape)				
Eastern Cape	-0.245*	0.434	-0.168	-0.074
	(0.132)	(0.289)	(0.130)	(0.064)
Northern Cape	0.048	0.379	0.185	0.043
	(0.110)	(0.238)	(0.114)	(0.058)
Free State	-0.100	-0.320	-0.206	0.050
	(0.146)	(0.242)	(0.156)	(0.085)
KwaZulu-Natal	-0.119	0.119	-0.144	0.045
	(0.140)	(0.229)	(0.131)	(0.064)
North West	-0.213	0.871***	-0.365**	0.029
	(0.211)	(0.236)	(0.173)	(0.079)
Gauteng	-0.005	-0.032	-0.270**	0.019
	(0.129)	(0.174)	(0.110)	(0.063)
Mpumalanga	-0.232	-0.076	-0.504***	0.024
	(0.156)	(0.213)	(0.130)	(0.067)
Limpopo	-0.491***	-0.022	-0.508***	0.130*
	(0.156)	(0.251)	(0.149)	(0.069)
Constant	4.433***	-1.723**	3.733***	-1.346***
	(0.886)	(0.769)	(0.285)	(0.118)
Exclusion restrictions				
Mother's occupation (base: Missing)				
Agri/Elementary			0.066	
			(0.053)	
Professionals			-0.172***	
			(0.065)	
Semi-skilled/operator			0.014	
			(0.132)	
Clerks/sales workers			-0.044	
			(0.066)	
Craft/trade			0.108	
			(0.143)	
Never worked			-0.058	
			(0.052)	
Father's occupation (base: Missing)				
Agri/Elementary			0.169**	

Continuation of Table A.2

	Transition eqn.		Initial	Retention
Base period covariate	Remain	Fall	conditions	eqn.
			(0.072)	
Professionals			-0.172***	
			(0.059)	
Semi-skilled/operator			0.129	
			(0.086)	
Clerks/sales workers			0.155**	
			(0.068)	
Craft/trade			0.068	
			(0.060)	
Never worked			0.009	
			(0.054)	
Mother's education (base: Incomplete secondary)				
Missing			-0.159*	
			(0.082)	
No schooling			-0.063	
			(0.071)	
Primary			-0.077	
			(0.076)	
Matric			-0.030	
			(0.072)	
Post-school qualification			-0.083	
			(0.064)	
Father's education (base: Incomplete secondary)				
Missing			0.016	
			(0.072)	
No schooling			0.004	
			(0.075)	
Primary			-0.018	
			(0.070)	
Matric			-0.165**	
			(0.078)	
Post-school qualification			-0.275***	
			(0.078)	
Respondent characteristics				
CSM or a TSM? (CSM=1)				1.128***
				(0.027)
Very friendly during interview (yes=1)				0.134***
				(0.041)
Very attentive during interview (yes=1)				0.022
				(0.034)
Observations	116 462		116 462	116 462

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. Standard errors are robust to arbitrary levels of intra-household correlation and the presence of repeated observations on the same individual. Data are weighted using post-stratification weights from the base period.

Table A.3 Average characteristics of households and household heads by funding class (youth respondents)

	Eligible		Missing Middle			
	Persistent	Transient	Vulnerable	Stable	Elite	Total
Weighted share of youth	60.90%	32.00%	1.97%	2.12%	3.01%	100%
(a) Characteristics of the head						
Household head is employed	0.39	0.61	0.65	0.84	0.78	0.49
Age of household head	44.13	42.42	49.13	40.95	43.27	43.59
Female household head	0.63	0.47	0.46	0.26	0.31	0.56
Race						
African	0.93	0.76	0.79	0.21	0.3	0.84
Coloured	0.06	0.12	0.13	0.13	0.09	0.08
Asian/Indian	0.01	0.03	0.06	0.08	0.1	0.02
White	0.00	0.10	0.02	0.58	0.50	0.06
Highest education						
Missing	0.00	0.01	0.00	0.01	0.01	0.00
No schooling	0.18	0.02	0.03	0.01	0.00	0.12
Primary	0.27	0.13	0.11	0.00	0.03	0.21
Incomplete secondary	0.41	0.22	0.33	0.05	0.09	0.33
Matric	0.09	0.22	0.2	0.17	0.18	0.14
Post-school qualification	0.04	0.40	0.35	0.76	0.69	0.20
(b) Household characteristics						
Income and expenditure						
Per capita expenditure[a]	972.52	2 989.42	5 026.34	10 163.72	13 230.04	2 261.99
Per capita income[a]	1 336.04	3 428.71	8 748.65	12 609.23	22 682.6	3 034.02
Share of income from[b]:						
Labour market	0.43	0.71	0.81	0.91	0.84	0.55
Government grants	0.34	0.12	0.05	0.01	0.01	0.25
Investment income	0.01	0.02	0.03	0.03	0.05	0.02
Remittances	0.08	0.04	0.04	0.02	0.02	0.07
Subsistence agriculture	0.01	0.00	0.00	0.00	0.00	0.00
One-shot response (no source)	0.12	0.09	0.08	0.04	0.08	0.11
Assets, credit, and infrastructure						
At least one resident has a:						
Home loan/bond	0.00	0.07	0.21	0.54	0.44	0.05
Bank loan	0.07	0.21	0.35	0.30	0.25	0.13
Study loan with a bank	0.00	0.02	0.05	0.04	0.04	0.01
Vehicle finance	0.00	0.08	0.28	0.40	0.41	0.05
Credit card	0.02	0.13	0.31	0.49	0.55	0.09
Store card	0.19	0.42	0.56	0.46	0.51	0.29
Financial assets	0.61	0.90	0.94	0.94	0.95	0.73
Pension assets	0.01	0.15	0.32	0.49	0.47	0.08
Household member owns dwelling	0.71	0.66	0.83	0.73	0.77	0.70
Household livestock assets	0.08	0.02	0.03	0.00	0.01	0.06
Number of rooms in house	3.64	4.80	5.58	5.84	6.54	4.18
Household has access to electricity	0.79	0.95	0.95	0.98	0.97	0.86
Piped water on site	0.66	0.92	0.93	0.99	0.96	0.76
Has a flush toilet	0.42	0.8	0.81	1.00	0.96	0.58
Household composition						
Number of household residents	5.19	5.25	6.04	3.70	4.22	5.17
Number of children under 6	0.84	0.63	0.84	0.35	0.34	0.74
Number of children aged 6–18	1.53	1.43	1.44	0.70	0.88	1.46
Number of elderly residents[c]	0.28	0.24	0.28	0.16	0.17	0.26
Household employment rate [d]	0.34	0.56	0.54	0.73	0.66	0.43
Location						
Traditional	0.43	0.20	0.21	0.04	0.06	0.34
Urban	0.51	0.77	0.76	0.94	0.93	0.62
Farm	0.06	0.02	0.03	0.02	0.02	0.05

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period.
[a] Monthly value in Rands. [b] Excludes imputed rental income. Labour market income reflects after tax income. [c] Aged 60+. [d] Of working age adults.

Table A.4 Average household characteristics, youth households vs. all households

	Youth	All	P-value
(a) Characteristics of the head			
Household head is employed	0.52	0.54	0.00
Age of household head	41.31	44.25	0.00
Household head is female	0.52	0.47	0.00
Race			
African	0.83	0.79	0.00
Coloured	0.08	0.08	0.18
Asian/Indian	0.02	0.02	0.46
White	0.07	0.11	0.00
Highest education			
Missing	0.01	0.00	0.41
No schooling	0.09	0.09	0.63
Primary	0.18	0.19	0.00
Incomplete secondary	0.34	0.32	0.00
Matric	0.16	0.16	0.03
Post-school qualification	0.23	0.24	0.00
(b) Household characteristics			
Income and expenditure			
Per capita expenditure	2 763.19	3 680.98	0.00
Per capita income	3 575.22	5 413.27	0.00
Share of income from[b]:			
Labour market	0.56	0.56	0.64
Government grants	0.22	0.21	0.00
Investment income	0.02	0.04	0.00
Remittances share	0.08	0.07	0.00
Subsistence agriculture	0.00	0.00	0.92
One-shot response (no source)	0.11	0.11	0.13
Assets, credit, and infrastructure			
At least one resident has a:			
Home loan/bond	0.06	0.07	0.00
Bank loan	0.12	0.12	0.78
Study loan with a bank	0.01	0.01	0.06
Vehicle finance	0.06	0.06	0.25
Credit card	0.09	0.10	0.00
Store card	0.27	0.24	0.00
Financial assets (indicator)	0.71	0.70	0.00
Pension assets (indicator)	0.09	0.10	0.00
Household member owns dwelling (indicator)	0.65	0.64	0.59
Household livestock assets (indicator)	0.04	0.04	0.00
Number of rooms in house	3.83	3.79	0.01
Access to electricity	0.85	0.85	0.21
Piped water on site	0.77	0.78	0.00
Has a flush toilet	0.61	0.64	0.00
Household composition			
Number of household residents	3.91	3.34	0.00
Number of children under 6	0.55	0.43	0.00
Number of children aged 6–18	1.06	0.85	0.00
Number of elderly residents (60+)[c]	0.21	0.26	0.00
Household employment rate (working age adults)[d]	0.48	0.52	0.00
Location			
Traditional	0.29	0.27	0.00
Urban	0.66	0.68	0.00
Farm	0.05	0.05	0.01

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period.
[a] Monthly value in Rands. [b] Excludes imputed rental income. Labour market income reflects after tax income. [c] Aged 60+. [d] Of working age adults.
Youth household is a household with at least one resident youth.

Table A.5 Average characteristics of households and household heads by threshold value and funding class (missing middle respondents)

	Threshold = 37.79%		Threshold = 50%		Threshold = 75%	
	Vulnerable	Stable	Vulnerable	Stable	Vulnerable	Stable
(a) Characteristics of the head						
Household head is employed	0.64	0.79	0.65	0.75	0.52	0.75
Age of household head	49.1	44.69	50.3	44.92	54.34	45.38
Female household head	0.48	0.26	0.51	0.29	0.59	0.32
Race						
African	0.76	0.16	0.83	0.24	0.94	0.35
Coloured	0.13	0.13	0.11	0.14	0.03	0.15
Asian/Indian	0.05	0.06	0.04	0.07	0.03	0.06
White	0.06	0.65	0.02	0.56	0.00	0.44
Highest education						
Missing	0.00	0.01	0.00	0.01	0.00	0.01
No schooling	0.03	0.01	0.03	0.01	0.05	0.01
Primary	0.09	0.00	0.13	0.00	0.26	0.01
Incomplete secondary	0.29	0.08	0.32	0.11	0.41	0.14
Matric	0.22	0.19	0.20	0.21	0.12	0.22
Post-school qualification	0.37	0.71	0.32	0.67	0.15	0.62
(b) Household characteristics						
Income and expenditure						
Per capita expenditure[a]	5 265.14	9 943.06	4 781.88	9 277.86	3 289.40	8 549.38
Per capita income[a]	8 835.43	11 986.31	8 218.88	11 691.68	6 233.05	11 283.60
Share of income from[b]:						
Labour market	0.80	0.88	0.79	0.88	0.69	0.88
Government grant	0.04	0.01	0.06	0.01	0.09	0.01
Other government income	0.01	0.00	0.02	0.00	0.03	0.00
Investment income	0.03	0.04	0.03	0.04	0.03	0.04
Remittances	0.03	0.01	0.03	0.01	0.05	0.01
Subsistence agriculture	0.00	0.00	0.00	0.00	0.01	0.00
One-shot response (no source)	0.08	0.05	0.08	0.06	0.10	0.06
Assets, credit, and infrastructure						
At least one resident has a:						
Home loan/bond	0.24	0.53	0.18	0.51	0.09	0.46
Bank loan	0.34	0.28	0.32	0.31	0.29	0.32
Study loan with a bank	0.03	0.03	0.04	0.02	0.05	0.03
Vehicle finance	0.30	0.35	0.27	0.35	0.16	0.36
Credit card	0.32	0.47	0.31	0.45	0.23	0.43
Store card	0.55	0.44	0.54	0.46	0.48	0.49
Financial assets	0.94	0.94	0.93	0.94	0.92	0.94
Pension assets	0.31	0.52	0.29	0.48	0.24	0.45
Household member owns dwelling	0.86	0.78	0.86	0.79	0.87	0.8
Household owns livestock assets	0.04	0.00	0.05	0.00	0.09	0.00
Number of rooms in house	5.47	5.93	5.58	5.78	5.48	5.76
Access to electricity	0.94	0.97	0.92	0.97	0.91	0.96
Piped water on site	0.94	0.99	0.92	0.99	0.87	0.98
Has a flush toilet	0.82	1.00	0.77	0.99	0.67	0.96
Household composition						
Number of household residents	5.84	3.78	6.34	3.92	8.28	4.10
Household dependency ratio	0.63	0.49	0.64	0.51	0.79	0.51
Number of children under 6	0.83	0.33	0.98	0.35	1.48	0.40
Number of children aged 6–18	1.56	0.87	1.61	0.98	2.14	1.02
Number of elderly residents [c]	0.28	0.26	0.33	0.24	0.46	0.24
Household employment rate[d]	0.57	0.74	0.54	0.72	0.48	0.69
Location						
Traditional	0.19	0.04	0.23	0.05	0.31	0.08
Urban	0.78	0.95	0.75	0.93	0.66	0.90
Farm	0.03	0.01	0.02	0.02	0.03	0.02

Source: Author's own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period.
[a] Monthly value in Rands. [b] Share excludes imputed rental income. Labour market income reflects after tax income.
[c] Aged 60+. [d] Of working age adults.

Table A.6 Multivariate probit transition equation, fitted to each pair of waves independently (coefficient estimates)

Base period covariate	Wave 1	Wave 2	Wave 3	Wave 4
Characteristics of the household head				
Race (base: White)				
African	1.027***	1.183***	0.857***	0.839***
	(0.175)	(0.200)	(0.196)	(0.184)
Coloured	0.586***	0.881***	0.585***	0.573***
	(0.226)	(0.228)	(0.210)	(0.205)
Asian/Indian	0.391	-0.094	0.288	0.367
	(0.339)	(0.325)	(0.310)	(0.287)
Age of household head	0.003	-0.008	0.005	-0.000
	(0.005)	(0.005)	(0.005)	(0.004)
Household head is employed (yes=1)	0.200	-0.097	0.474***	0.023
	(0.137)	(0.155)	(0.149)	(0.111)
Household head is female (yes=1)	0.024	0.174	0.221**	0.279***
	(0.099)	(0.122)	(0.096)	(0.081)
Highest education (base: Incomplete secondary)				
Missing[a]	-0.928*	-	-0.967***	0.505
	(0.479)	-	(0.314)	(0.412)
No schooling	0.567*	0.447	-0.058	-0.158
	(0.314)	(0.298)	(0.205)	(0.206)
Primary	0.156	0.510***	-0.010	-0.316**
	(0.172)	(0.166)	(0.165)	(0.155)
Matric	-0.701***	-0.270	-0.481***	0.050
	(0.161)	(0.181)	(0.148)	(0.172)
Post-school qualification	-0.706***	-0.643***	-0.660***	-0.767***
	(0.166)	(0.151)	(0.153)	(0.119)
Household characteristics				
Household employment rate	-0.594***	-0.663***	-0.837***	-0.521***
	(0.179)	(0.217)	(0.180)	(0.153)
At least one resident has financial assets (yes=1)	0.024	-0.079	0.043	-0.133
	(0.130)	(0.148)	(0.138)	(0.138)
At least one resident has pension assets (yes=1)	-0.201	-0.541***	-0.505***	-0.523***
	(0.173)	(0.136)	(0.175)	(0.109)
Household member owns dwelling (yes=1)	0.020	-0.308*	-0.212	-0.360**
	(0.155)	(0.163)	(0.134)	(0.144)
Household has livestock assets (yes=1)	0.526*	0.290	0.090	0.661***
	(0.276)	(0.270)	(0.187)	(0.220)
Share of durables owned	-1.420***	-0.123	-2.565***	-3.438***
	(0.538)	(0.527)	(0.489)	(0.461)
Number of rooms in house	-0.060***	-0.082***	-0.061***	-0.050***
	(0.023)	(0.026)	(0.020)	(0.019)
Household has access to electricity (yes=1)	0.016	-0.107	0.061	0.162
	(0.148)	(0.183)	(0.198)	(0.183)
Piped water on site (yes=1)	0.073	-0.003	-0.004	0.216
	(0.205)	(0.219)	(0.209)	(0.160)
Household has a flush toilet (yes=1)	0.074	-0.705***	-0.432*	0.150
	(0.248)	(0.173)	(0.236)	(0.148)
Number of household residents	-0.179***	0.001	-0.020	0.025
	(0.058)	(0.053)	(0.038)	(0.034)
Household dependency ratio	-0.125	-0.072	0.343***	0.253***
	(0.086)	(0.098)	(0.114)	(0.074)

Continuation of Table A.6

Base period covariate	Wave 1	Wave 2	Wave 3	Wave 4
Number of children under 6	0.305***	0.166	-0.026	-0.166**
	(0.101)	(0.117)	(0.090)	(0.066)
Number of children aged 6-18	0.203***	-0.080	-0.088	-0.092
	(0.068)	(0.082)	(0.064)	(0.065)
Location (base: Rural)				
Urban	-0.061	0.062	-0.123	-0.336*
	(0.255)	(0.179)	(0.266)	(0.173)
Farm	0.533*	0.606*	0.261	0.247
	(0.320)	(0.321)	(0.404)	(0.277)
Province & time fixed effects	Yes	Yes	Yes	Yes
Pseudo log-likelihood	-27 761 214	-29 616 459	-29 246 099	-32 647 150
Model chi-squared (d.f.=129)	1 169.42	1 613.40	2 359.26	2 539.10
Number of clusters	5 908	8 275	10 662	13 472
Observations	22 584	27 219	31 288	35 371

Source: Author's own calculations using NIDS Waves 1 to 5.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. Standard errors are robust to arbitrary levels of intra-household correlation and the presence of repeated observations on the same individual. Data are weighted using post-stratification weights from the base period.
Column title reflects the base wave for each transition.
[a] Household head education missing is dropped in the Wave 2/Wave 3 transition because of collinearity (it predicts success perfectly). There are only 48 observations in the estimation sample of the Wave 2/Wave 3 transition that had household head education missing.

Table A.7 Average characteristics of households and household heads (wave 1-2 transition)

	Eligible		Missing Middle			
	Persistent	Transient	Vulnerable	Stable	Elite	**Total**
Share of respondents	55.76%	36.31%	1.94%	1.73%	4.26%	100%
(a) Characteristics of the head						
Household head is employed	0.39	0.55	0.74	0.76	0.62	0.47
Age of household head	49.64	45.21	48.47	45.43	46.35	47.80
Female household head	0.53	0.40	0.22	0.14	0.11	0.45
Race						
African	0.96	0.67	0.70	0.06	0.20	0.80
Coloured	0.03	0.17	0.12	0.1	0.06	0.09
Asian/Indian	0.01	0.04	0.12	0.04	0.11	0.02
White	0.00	0.12	0.06	0.79	0.63	0.09
Highest education						
Missing	0.00	0.02	0.00	0.00	0.01	0.01
No schooling	0.30	0.04	0.00	0.00	0.01	0.18
Primary	0.39	0.17	0.05	0.00	0.02	0.28
Incomplete secondary	0.29	0.26	0.41	0.11	0.05	0.27
Matric	0.02	0.26	0.11	0.35	0.18	0.12
Post-school qualification	0.01	0.26	0.43	0.54	0.72	0.14
(b) Household characteristics						
Income and expenditure						
Per capita expenditure[a]	765.44	2 792.27	7 096.5	9 862.29	15 338.88	2 401.93
Per capita income[a]	888.87	2 765.53	9 379.95	11 121.68	25 392.68	2 954.99
Share of income from[b]:						
Labour market	0.39	0.74	0.85	0.95	0.87	0.56
Government grant	0.47	0.16	0.01	0.01	0.01	0.32
Other government income	0.01	0.01	0.07	0.00	0.00	0.01
Investment income	0.01	0.03	0.05	0.05	0.06	0.02
Remittances	0.06	0.05	0.02	0.00	0.05	0.05
Subsistence agriculture	0.01	0.00	0.00	0.00	0.00	0.01
One-shot response (no source)	0.03	0.01	0.00	0.00	0.00	0.02
Assets, credit, and infrastructure						
At least one resident has a:						
Home loan/bond	0.01	0.12	0.29	0.66	0.50	0.09
Bank loan	0.04	0.16	0.38	0.26	0.15	0.10
Study loan with a bank	0.00	0.02	0.10	0.01	0.00	0.01
Vehicle finance	0.01	0.10	0.47	0.34	0.49	0.08
Credit card	0.02	0.16	0.45	0.61	0.67	0.12
Store card	0.12	0.36	0.68	0.51	0.44	0.24
Financial assets	0.4	0.75	0.99	0.99	0.96	0.57
Pension assets	0.01	0.11	0.27	0.51	0.39	0.08
Household member owns dwelling	0.82	0.70	0.83	0.76	0.88	0.78
Household owns livestock assets	0.16	0.03	0.00	0.00	0.02	0.10
Number of rooms in house	3.73	4.73	4.93	6.28	7.31	4.30
Access to electricity	0.69	0.9	0.97	0.93	0.98	0.79
Piped water on site	0.55	0.88	0.98	1.00	0.96	0.7
Has a flush toilet	0.33	0.74	0.95	0.99	0.95	0.53
Household composition						
Number of household residents	5.72	5.24	4.99	3.85	4.02	5.42
Household dependency ratio	0.96	0.66	0.63	0.51	0.57	0.82
Number of children under 6	1.02	0.68	0.45	0.41	0.40	0.85
Number of children aged 6–18	1.97	1.48	1.58	0.71	1.00	1.72
Number of elderly residents[c]	0.35	0.28	0.19	0.15	0.17	0.31
Household employment rate [d]	0.29	0.48	0.55	0.66	0.59	0.38
Location						
Traditional	0.50	0.20	0.09	0.02	0.05	0.36
Urban	0.42	0.77	0.89	0.97	0.94	0.59
Farm	0.08	0.03	0.02	0.01	0.01	0.05

Source: Author's own calculations using NIDS Waves 1 to 2 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period. [a] Monthly value in Rands. [b] Share excludes imputed rental income. Labour market income reflects after tax income. [c] Aged 60+. [d] Of working age adults.

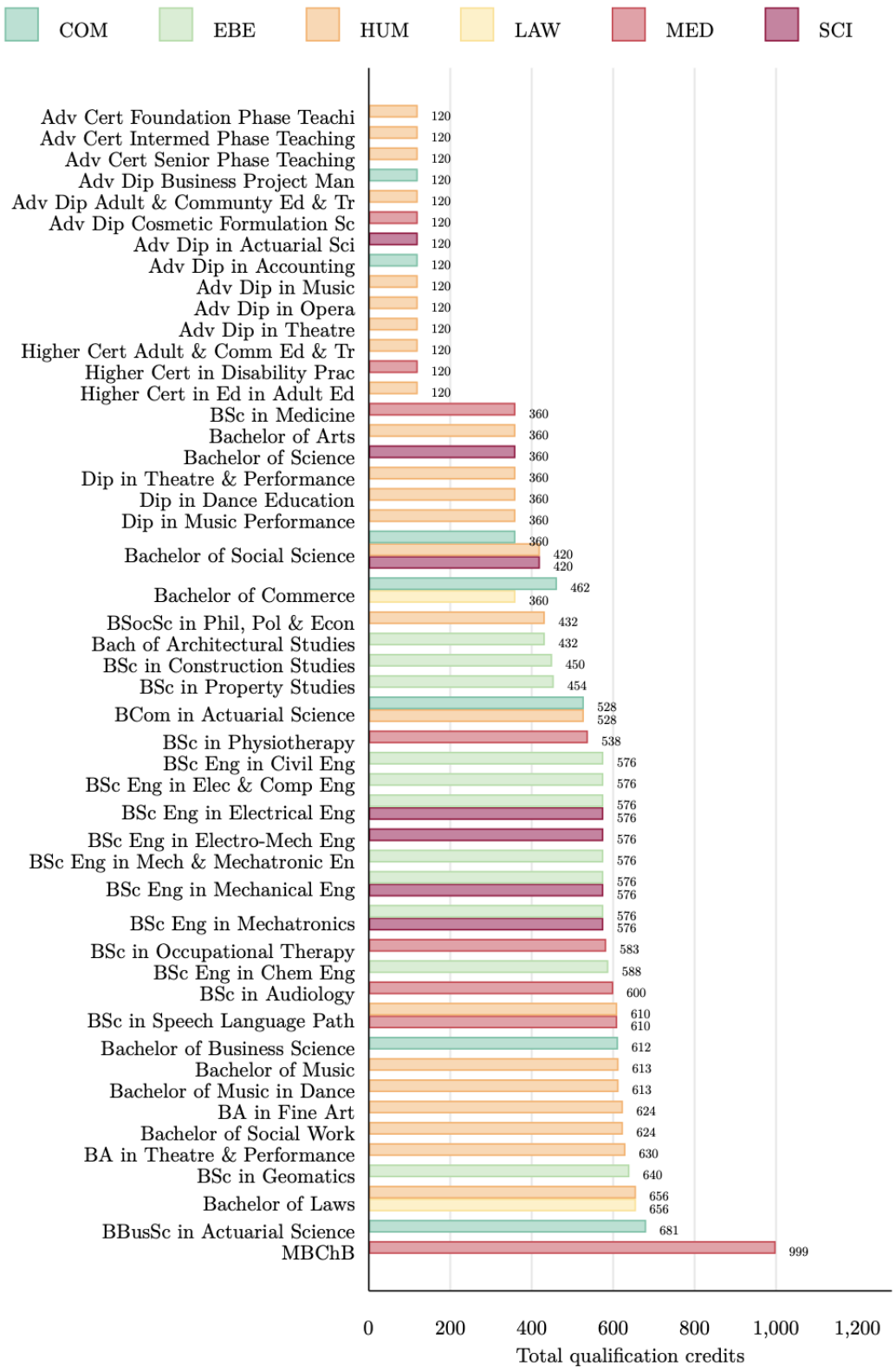
Appendix B

Table B.1 Undergraduate qualification credits by minimum time for degree completion

Minimum time (years)	One	Two	Three	Four	Six
(a) Qualifications					
Number of qualifications offered	13	1	22	36	2
Number of qualifications by faculty					
Commerce	2	0	4	5	0
Engineering and the Built Environment	0	0	3	12	0
Humanities	8	1	9	7	0
Law	0	0	2	1	0
Health Sciences	2	0	1	7	2
Science	1	0	3	4	0
(b) Qualification credits					
Average credits	120	120	418	601	999
Average credits per year	120	60	140	150	167
Average credits by faculty					
Commerce	120	-	453	640	-
Engineering and the Built Environment	-		445	589	-
Humanities	120	120	393	624	-
Law	-		508	656	-
Health Sciences	120	-	360	579	999
Science	120	-	380	576	

Source: Author's own calculations using UCT (2016-2021) and HEMIS (2012-2020).
Notes: One observation per qualification.

Figure B.1 Minimum qualification credits by qualification and faculty



Source: Author's own calculations using UCT (2016-2021) and HEMIS (2012-2020).

Table B.2 DD regression estimates, interactions with GPA quartile and NSFAS status

	2020 performance			2021 performance		
	Taken	Passed	Ratio	Taken	Passed	Ratio
Quartile 2	36.817***	53.833***	0.306***	35.503***	48.692***	0.282***
	(1.322)	(1.380)	(0.009)	(1.413)	(1.538)	(0.009)
Quartile 3	42.734***	70.600***	0.401***	41.198***	64.742***	0.374***
	(1.265)	(1.300)	(0.008)	(1.346)	(1.444)	(0.009)
Quartile 4	44.226***	77.473***	0.435***	44.449***	73.576***	0.412***
	(1.250)	(1.261)	(0.008)	(1.327)	(1.396)	(0.008)
NSFAS	6.411***	-1.659	-0.003	5.788***	-3.002	-0.013
	(1.772)	(1.764)	(0.012)	(1.901)	(1.983)	(0.013)
NSFAS·Quartile 2	-7.377***	-5.438**	-0.021	-6.815***	-5.810**	-0.020
	(2.151)	(2.371)	(0.015)	(2.305)	(2.636)	(0.016)
NSFAS·Quartile 3	-8.144***	-3.979*	-0.010	-7.312***	-3.087	-0.003
	(2.132)	(2.343)	(0.014)	(2.246)	(2.552)	(0.015)
NSFAS·Quartile 4	-7.043***	-0.623	0.007	-7.963***	-0.646	0.015
	(2.302)	(2.434)	(0.014)	(2.490)	(2.753)	(0.015)
Pandemic effect	-1.327	13.037***	0.095***	23.932***	8.565***	0.037**
	(2.048)	(1.962)	(0.013)	(2.372)	(2.685)	(0.017)
Pandemic effect·NSFAS	-0.529	-3.489	-0.007	-6.126*	-13.208***	-0.084***
	(2.914)	(2.829)	(0.019)	(3.163)	(3.686)	(0.024)
Pandemic effect·Quartile 2	-1.770	-7.787***	-0.036**	-22.698***	-12.690***	-0.056***
	(2.459)	(2.454)	(0.015)	(2.836)	(3.326)	(0.020)
Pandemic effect·Quartile 3	0.956	-10.382***	-0.075***	-21.201***	-12.422***	-0.074***
	(2.317)	(2.275)	(0.014)	(2.699)	(3.126)	(0.018)
Pandemic effect·Quartile 4	3.951*	-10.251***	-0.080***	-20.875***	-8.620***	-0.046***
	(2.231)	(2.171)	(0.013)	(2.610)	(2.930)	(0.017)
Pandemic effect·NSFAS·Quartile 2	0.524	2.321	-0.007	4.099	7.309	0.036
	(3.586)	(3.738)	(0.023)	(3.981)	(4.916)	(0.030)
Pandemic effect·NSFAS·Quartile 3	-0.900	1.743	0.010	-1.922	5.462	0.055*
	(3.483)	(3.609)	(0.022)	(3.935)	(4.781)	(0.028)
Pandemic effect·NSFAS·Quartile 4	-5.522	-3.922	-0.017	2.290	4.175	0.036
	(3.722)	(3.827)	(0.022)	(4.219)	(5.119)	(0.028)
3rd year	8.053***	9.941***	0.039***	7.133***	7.931***	0.026***
	(0.512)	(0.524)	(0.003)	(0.595)	(0.585)	(0.003)
4th year	-7.498***	-3.076***	0.037***	-4.188***	2.077**	0.061***
	(0.786)	(0.788)	(0.004)	(0.879)	(0.939)	(0.005)
Constant	106.264***	69.203***	0.499***	106.666***	72.841***	0.518***
	(1.124)	(1.082)	(0.007)	(1.208)	(1.213)	(0.008)
R^2	0.13	0.28	0.26	0.11	0.24	0.24
Observations	32 585	32 585	32 581	28 713	28 713	28 711

Source: Author's own calculations using UCT (2016-2021).
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses and are robust to the presence of repeated observations on the same individual. Sample comprises those in entrance cohorts 2016-2019.

Table B.3 DD regression estimates: 2021 performance with indicator for improvement in prior year

	Ratio
2021 effect	-0.244***
	(0.019)
$Increase_{t-1}$	0.140***
	(0.010)
2021 effect·$Increase_{t-1}$	0.121***
	(0.024)
Quartile 2	0.203***
	(0.008)
Quartile 3	0.297***
	(0.007)
Quartile 4	0.341***
	(0.007)
2021 effect·Quartile 2	0.194***
	(0.022)
2021 effect·Quartile 3	0.203***
	(0.020)
2021 effect·Quartile 4	0.225***
	(0.020)
$Increase_{t-1}$·Quartile 2	-0.074***
	(0.012)
$Increase_{t-1}$·Quartile 3	-0.152***
	(0.017)
$Increase_{t-1}$·Quartile 4	-0.182***
	(0.028)
2021 effect·$Increase_{t-1}$·Quartile 2	-0.163***
	(0.029)
2021 effect·$Increase_{t-1}$·Quartile 3	-0.114***
	(0.034)
2021 effect·$Increase_{t-1}$·Quartile 4	-0.142**
	(0.068)
3rd year	-0.009***
	(0.003)
4th year	0.025***
	(0.004)
Constant	0.632***
	(0.006)
R^2	0.23
Observations	26 687

Source: Authors' own calculations using UCT (2016-2021).
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses and are robust to the presence of repeated observations on the same individual. Sample comprises those in entrance cohorts 2016-2019. Dropouts excluded. The dependent variable is the share of credits passed.

Appendix C

Table C.1 Tobit average marginal effects

Variables	Probability eqn.	Amount eqn.
Education (base: incomplete secondary)		
No schooling	-0.015	-41.350
	(0.018)	(50.490)
Primary school	-0.002	-4.730
	(0.014)	(37.249)
Matric	0.005	12.937
	(0.011)	(29.772)
Qualification not requiring matric	0.026	69.874
	(0.016)	(43.512)
Post-secondary	0.048***	117.923***
	(0.015)	(41.324)
Age (years)	0.001***	3.952***
	(0.000)	(1.233)
Male (yes=1)	0.044***	118.308***
	(0.011)	(28.106)
Scaled labour market income	0.005***	13.848***
	(0.001)	(2.991)
Scaled non-labour market income	0.003	8.173
	(0.002)	(5.732)
Wage employment (yes=1)	0.140***	358.328***
	(0.014)	(38.154)
Casual/self-employment (yes=1)	0.077***	205.466***
	(0.016)	(45.306)
Number of children aged 6-16 in the household	-0.033***	-89.955***
	(0.005)	(14.054)
Number of children under 6 in the household	-0.026***	-70.101***
	(0.005)	(15.381)
Co-resident mother (yes=1)	-0.059***	-164.236***
	(0.009)	(26.538)
Co-resident father (yes=1)	-0.035**	-96.517**
	(0.015)	(44.825)
Rural location (yes=1)	-0.005	-13.388
	(0.011)	(30.947)
Observations	18 913	18 913

Source: Authors' own calculations using NIDS Wave 5.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. Standard errors are robust to clustering at the cluster level. Data are weighted using post-stratification weights. The AMEs in the first column reflect the effects of covariates on the probability of remitting (a positive value observed). The AMEs in the second column reflect the effect of the covariates on the value of cash and in-kind remittances sent in the last month, conditional on a positive value being sent.

Table C.2 The lognormal hurdle model (amount equation)

Variables	Coefficient estimates
Education (omitted: incomplete secondary)	
No schooling	0.097
	(0.152)
Primary school	0.020
	(0.143)
Matric	-0.013
	(0.097)
Qualification not requiring matric	-0.013
	(0.092)
Post-secondary	0.272**
	(0.114)
Age (years)	0.058***
	(0.015)
Age squared	-0.001***
	(0.000)
Male (yes=1)	0.192***
	(0.070)
Scaled labour market income	0.032***
	(0.012)
Scaled non-labour market income	0.016
	(0.016)
Graduate & labour market income interaction	-0.011
	(0.013)
Graduate & non-labour market income interaction	-0.026
	(0.024)
Wage employment (yes=1)	0.176
	(0.113)
Casual/self-employment (yes=1)	0.012
	(0.134)
Number of children aged 6-16 in the household	-0.089**
	(0.035)
Number of children under 6 in the household	-0.065
	(0.045)
Co-resident mother (yes=1)	-0.004
	(0.130)
Co-resident father (yes=1)	0.012
	(0.139)
Rural location (yes=1)	-0.145*
	(0.080)
Constant	5.132***
	(0.314)
R^2	0.20
Observations	1 664

Source: Author's own calculations using NIDS Wave 5.
Notes: OLS regression on the logged value of positive cash and in-kind remittances sent in the last month. *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. Standard errors are robust to clustering at the cluster level. Data are weighted using post-stratification weights.