Graduate unemployment and Higher Education Institutions in South Africa

HENDRIK VAN BROEKHUIZEN

Stellenbosch Economic Working Papers: 08/16

KEYWORDS: GRADUATES, UNEMPLOYMENT, HIGHER EDUCATION
JEL: J64, I23, I26

HENDRIK VAN BROEKHUIZEN
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: HENDRIKVANB@SUN.AC.ZA

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

BER
BUREAU FOR ECONOMIC RESEARCH

A WORKING PAPER OF THE DEPARTMENT OF ECONOMICS AND THE
BUREAU FOR ECONOMIC RESEARCH AT THE UNIVERSITY OF STELLENBOSCH

Graduate unemployment and Higher Education Institutions in South Africa

HENDRIK VAN BROEKHUIZEN

ABSTRACT

The emerging consensus regarding high and rising levels of graduate unemployment in South Africa in recent years has primarily been based on a select number of studies, all of which have serious shortcomings ranging from deficient definitions of "graduates" to the use of outdated, incomplete, or unrepresentative data. Moreover, given the heterogeneity in the quality of higher education in South Africa, existing findings regarding aggregate graduate unemployment in the country, even if accurate, mask the substantial variation in labour market outcomes which are likely to be faced by graduates from different higher education institutions. This paper attempts to address these issues by examining graduate unemployment and employment in South Africa with specific emphasis on the type and quality of higher education using multiple labour force survey and administrative datasets. Its primary contribution is to incorporate the effect of potential measures of higher education institution type and quality on the likelihood of graduate unemployment and employment by probabilistically linking graduates that are observed in labour force survey data to the institutions from which they are likely to have graduated given their time-invariant observable characteristics and the known demographic composition of the graduates produced by each of South Africa's formal higher education institutions every year. The analysis shows that graduate unemployment in South Africa is not only low in relation to overall unemployment in the country, but that much of the racially-delineated differentials in graduate unemployment and employment outcomes can likely be attributed to heterogeneity in the quality and type of higher education institutions commonly attended by individuals from different racial backgrounds.

Keywords: graduates, unemployment, higher education
JEL codes: J64, I23, I26

1 Introduction

Since the early 2000s, new microeconomic research has increasingly suggested that the relative labour market benefits of Higher Education (HE) in South Africa may be on the decline.[1] The apparent significant rise in graduate unemployment rates between 1995 and 2005 and the extent of emerging skills-mismatches, according to which the skills that new graduate labour market entrants possess deviate from the skills that employers demand, are two areas that have received much attention, both in academic research and the media (Koen, 2006; Branson *et al.*, 2009*b*:2).

The supposed deterioration of graduate labour market outcomes in South Africa is often attributed to a combination of the HE system's lack of responsiveness to structural changes in the domestic economy since 1994 and changes in the underlying demographic composition of South Africa's pool of graduate labour force participants and the fields in which they chose to study (Bhorat, 2004; DRPU, 2006; Pauw *et al.*, 2008). In a review of the South African literature on unemployment among individuals with post-secondary qualifications, Kraak (2010) argues that this skills-mismatch has exacerbated South Africa's existing skills shortages and adversely affected the employability and subsequent labour market prospects faced by tertiary-educated individuals to a greater extent than for any other educational cohort.

Despite frequent references in the media and political statements to worsening labour market outcomes for South African graduates, the shortcomings of existing research on the relationship between HE and the labour market imply that there is still much confusion about the labour market prospects that graduates are likely to face. This confusion is exacerbated by prominent reporting of graduate employment and unemployment figures that are outdated, unverified, or taken out of context. Furthermore, it is still not well-understood why there appear to be persistent differentials in the labour market outcomes for graduates from different race groups, or how the specific higher education institutions (HEIs) that graduates attend relate to their expected labour market outcomes.

This paper aims to provide clarity on some hitherto unanswered questions regarding graduate labour market outcomes by examining the relationship between HEIs and the probability of unemployment and employment in the South African labour market. By focussing on both the probability of employment and unemployment, the research aims to firstly assess the scale and scope of South Africa's apparent graduate unemployment problem in the context of other developments that have affected the domestic labour market and the HE system over time. The objective of the multivariate analysis is not only to estimate the magnitude of the labour market premiums associated with participation in HE in terms of lowering the likelihood of unemployment and raising the likelihood of employment in South Africa, but to also incorporate the effects of HEI type on employment and unemployment outcomes by probabilistically linking graduates to the known distributions of annual graduate outputs from the public HE system, based on time-invariant demographic characteristics.[2]

The results from the analysis reveal that graduate unemployment in South Africa is not rising significantly over time and that it is, in fact, low in relation to overall unemployment in the country. Given the significant

[1] See, for example, Bhorat (2004:957 - 961), DRPU (2006), Scott *et al.* (2007:5), Altman (2007:11), Pauw *et al.* (2008), Kraak (2010), Maharasoa and Hay (2010), Van der Merwe (2010), Naong (2011), NPC (2011:317), Bhorat and Mayet (2012:30 - 31), Bhorat *et al.* (2010), CHEC (2013:7 - 10), Baldry (2015), and Kraak (2015).

[2] The data on South Africa's private HE sector is hihgly fragmented, but recent estimates suggest that it accounts for only a negligible percentage of all HE graduate outputs in South Africa Blom (2011); DHET (2015). It is therefore excluded from the discussion and analysis in this paper.

changes that have occurred in South Africa's HE system over the past 25 years, the results from the multivariate analysis show that much of the unexplained differences in employment and unemployment rates between Black, Coloured, Indian, and White graduates may be attributed to differences in the types of HEIs that different race groups have historically been likely to attend. These findings suggest that graduate unemployment in the country is not a general problem and that interventions aimed at improving the employment prospects of historically disadvantaged graduates should be targeted at improving the functionality of historically disadvantaged HEIs, rather than entailing wide-scale reform of South Africa's HE system as a whole.

2 The literature on graduate unemployment and employment in South Africa

Despite the limited attention that has historically been given to graduate labour market outcomes and their potential implications in the context of South Africa's broader labour market challenges, a number of prominent studies released since 2000 have raised concerns that graduate unemployment may rapidly be emerging as a significant problem in the country. In one of the earliest of these studies, Bhorat (2004), using data from the 1995 October Household Survey (OHS) and March 2002 Labour Force Survey (LFS), finds that, amidst rising overall unemployment rates, the broad unemployment rate for tertiary-educated individuals increased by 139% between 1995 and 2002 – by far the largest increase in unemployment for any education cohort. More worrying, however, is the fact that these rises in unemployment rates appeared to have been greatest for individuals with degrees and post-graduate qualifications, with White and Black graduate broad unemployment rates rising by 141% and 280%, respectively, over the 7-year period (Bhorat, 2004:959).

Bhorat (2004)'s substantive findings have received support in a number of papers published since 2004. Notable among these are the studies by DRPU (2006) and later Pauw *et al.* (2008) and Kraak (2010). The results from the descriptive analysis by DRPU (2006) showed that the increase in broad unemployment rates for tertiary-educated individuals from 6.6% in 1995 to 9.7% in 2005 was the largest for all education groups, despite levels of tertiary unemployment remaining low in relative terms (DRPU, 2006:8). The DPRU report also showed that graduate employment and unemployment rates varied substantially across race groups, suggesting that higher levels of unemployment among Black graduates, in particular, could at least partly be ascribed to the poor quality (or the perceived poor quality) of many HEIs in conjunction with the poor performance of the majority of the historically disadvantaged formal schooling system (DRPU, 2006:18-20). In other words, the extent of heterogeneity in the quality of HEIs may have eroded employer confidence in the productivity-signalling effect of HE qualifications, resulting in a shift in demand towards more experienced rather than more qualified employees (DRPU, 2006:21).

The finding that the employability of South Africa's HE-educated individuals, when measured in terms of the probability of being employed rather than unemployed, varies substantially by race has been emphasised in a large number of papers, most of which have relied on descriptive analyses and the use of nationally representative labour force data sources to draw inferences about changes in the employment and unemployment patterns for tertiary-educated individuals over time.[3]

More recent studies have also sought to identify the impact that HEI type and quality have on graduate employment and unemployment probabilities. Using data on seven South African universities from the Human

[3] See, for example, Mlatsheni and Rospabe (2002), Kruss (2007:683), Pauw *et al.* (2008:49 - 53), Branson *et al.* (2009*a*), Maharasoa and Hay (2010:141 - 142), Kraak (2010), Moleke (2010:89 - 92), Fisher and Scott (2011) and Bhorat *et al.* (2010).

Sciences Research Council's (HSRC) Graduate Destination Study, Bhorat *et al.* (2010) find that graduates who attended historically disadvantaged institutions (HDI) have significantly poorer labour market prospects than graduates from historically advantaged institutions (HAI), both in terms of initial absorption into employment and the ultimate incidence of unemployment. Similarly, Branson *et al.* (2009*a*) use data from the Cape Area Panel Study (CAPS) and find that the type of HEI at which individuals in the Western Cape province complete their tertiary studies has a significant impact on the labour market outcomes which they subsequently face.

2.1 Criticisms of the existing literature on graduate employment and unemployment in South Africa

The substantive conclusions drawn from studies noting adverse changes in the labour market prospects faced by graduates in South Africa resonate with those from international studies which have suggested that structural changes in other labour markets around the world have lead to a global trend of worsening labour market prospects for individuals with HE qualifications.[4] Consequently, the nature of the link between participation in HE and expected labour market outcomes is increasingly coming under question, both in South Africa and abroad. However, the majority of studies that have been conducted for the domestic labour market share common methodological shortcomings which mean that their findings are subject to a number of caveats.

First, few studies adequately differentiate between individuals with university degrees and individuals with post-secondary certificates and/or diplomas when analysing and drawing conclusions about the labour market prospects of the tertiary-educated, despite the fact that the two groups have been shown to differ vastly in terms of expected labour market outcomes (Koen, 2006:21). As shown in Section 3 below, this leads to a significant upward-biased perception of graduate unemployment and worsening graduate labour market prospects in the country.

Second, there is a tendency to draw causal inferences about the relationship between HE and labour market outcomes and strong conclusions about aggregate trends in the labour market outcomes for tertiaries from descriptive analyses conducted on data which is either not representative (Branson *et al.*, 2009*a*; Bhorat *et al.*, 2010; CHEC, 2013; Baldry, 2015), incomplete (Bhorat, 2004; DRPU, 2006; Pauw *et al.*, 2008), or dated (Pauw *et al.*, 2008; Kraak, 2010). Moreover, according to Yu (2008, 2010), there is good reason to doubt the accuracy of labour market outcome information for tertiary-educated respondents in the 1995 October Household Survey - the dataset which many of the most prominent studies of tertiary labour market outcomes in South Africa have used as the reference point for their empirical analyses.

Third, few studies sufficiently emphasise the levels of uncertainty that underlie their empirical methodologies and the confidence intervals which surround their reported point estimates, despite the fact that the sample sizes on which those estimates are based are often very small and that confidence intervals are therefore likely to be large. Rarely is any attempt made to establish the statistical significance of the differences between relevant point estimates when drawing conclusions regarding the trends in, and levels of, labour market outcomes for graduates. Instead, the significance of such "trends" appear to be inferred simply by comparing the inter-temporal changes in labour market outcome point estimates for individuals with HE qualifications with those for other education cohorts.

[4] See, for example, Teichler (2007), Nunez and Livanos (2010), (Wu, 2011), and (Humburg *et al.*, 2012).

Finally, with the exception of more recent studies like those by Branson *et al.* (2009*a*), Moleke (2010), Bhorat *et al.* (2010), CHEC (2013) and Baldry (2015), limited attention has thus far been given to the importance of heterogeneous HEI quality and historical patterns of access to HEIs in explaining racial labour market outcome differentials in South Africa, despite the fact that most studies find substantial differences in the employment and unemployment rates for tertiaries from different race groups. Consequently, little is known about the extent to which HE institutional considerations shape the labour market prospects of South African graduates.[5] Yet, in order to understand the nature of racial labour market outcome differentials and the potential causal mechanisms that drive them, it is necessary to take changes in South Africa's HE landscape and the demographic composition of its stock of graduates over time into account.

3 The South African graduate labour market

To understand the pitfalls of analysing the labour market outcomes for all tertiary-educated individuals as though they constitute an homogeneous group of individuals and referring to them as "graduates", it is necessary to illustrate the marked differences in labour market status outcomes for those individuals with diplomas and/or certificates from either TVET colleges or HEIs and individuals with university degrees obtained exclusively from HEIs. The former group is hereafter collectively referred to as *diplomates* and comprises all HE- or TVET-educated individuals with National Qualification Framework (NQF) exit level 5 or 6 qualifications. By contrast, the latter group is hereafter collectively referred to as *graduates*, comprising all HE-educated individuals with NQF exit level 7 or higher qualifications. The breakdown of the types of qualifications that are currently and have historically been awarded by South Africa's HEI along with their associated NQF exit level classifications is presented in Table A.1 in Appendix Appendix A.

Figure 1 shows the sizes of the narrow labour force and magnitudes of the narrow labour force participation (LFP) rates for graduates and diplomates over the period 2000 - 2015.[6] Other than the fact that diplomates represent a larger share of the tertiary-educated labour force than graduates, the two groups appear to have remarkable similar narrow LFP rates. However, the differences between these two groups become evident when looking at their respective employment rates as shown in Figure 2.

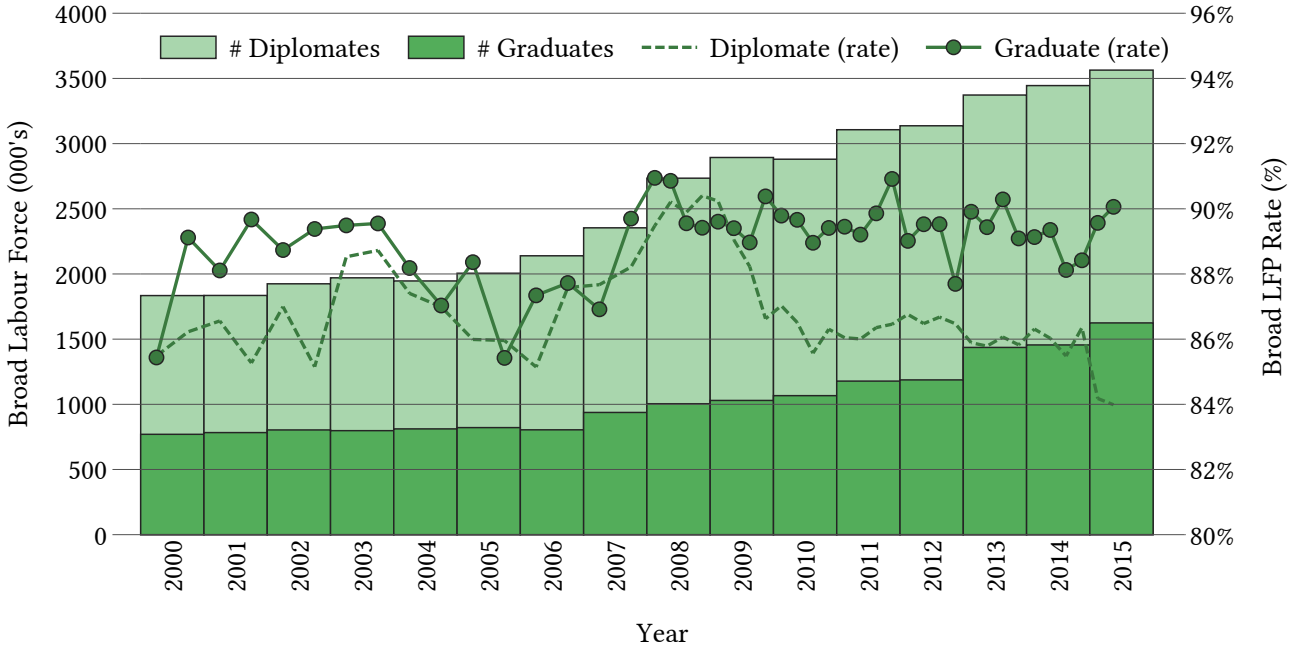
While diplomates still account for a larger share of employment among the tertiary-educated than graduates do, the graduate employment rate has consistently been between 5 and 10 percentage points higher than the diplomate employment rate over the period under consideration. The differences between the two tertiary-educated groups become even clearer when looking at narrow unemployment and narrow unemployment rates over the period.

Figure 3 shows that diplomates dominate narrow unemployment among the tertiary-educated. By 2015, more than 322 000 of the roughly 423 000 narrowly-unemployed tertiary-educated held diplomate-level qualifications. By contrast, not once since 2000 have graduates constituted more than 27% of the tertiary-educated

[5] Bhorat *et al.* (2010) is perhaps the only major recent study that has attempted to ascertain the impact of HEI quality on labour market outcomes in South Africa. Unfortunately, while the HSRC Graduate Destination data on which their analysis is based may be uniquely detailed, it is also inherently unrepresentative. Their results and conclusions are therefore unlikely to be reflective of the graduate labour market experience at a national level.

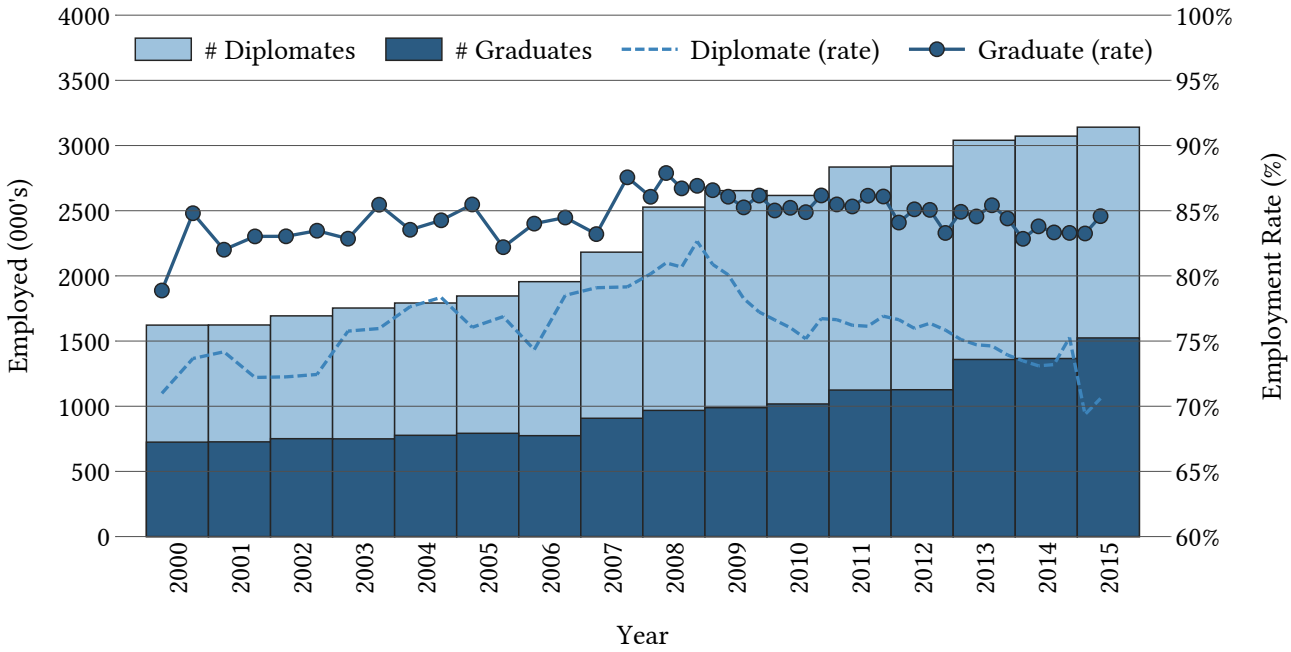
[6] Unless explicitly stated otherwise, the narrow definition of the labour force is used throughout this paper as it is the most consistent definition used across the various Labour Force and Quarterly Labour Force survey datasets that are used in the descriptive and multivariate analysis below. The narrow labour force is defined as all employed individuals plus all individuals who are not employed, but are either actively seeking employment or are planning on returning to existing jobs or enterprises soon.

Figure 1: Narrow labour force and narrow LFP rates (%) for graduates and diplomates (2000 - 2015)



NOTES: Own estimations using Stats SA's March 2000 LFS - QLFS2015Q2 data. Estimates are weighted and are calculated only for graduates and diplomates in the population of working-age (15 - 64 year-olds). Bars denote the respective sizes of the graduate and diplomate labour forces and have been stacked (measured on the left-hand-side vertical axis). Lines denote the respective graduate and diplomate LFP rates (measured on the right-hand-side vertical axis).

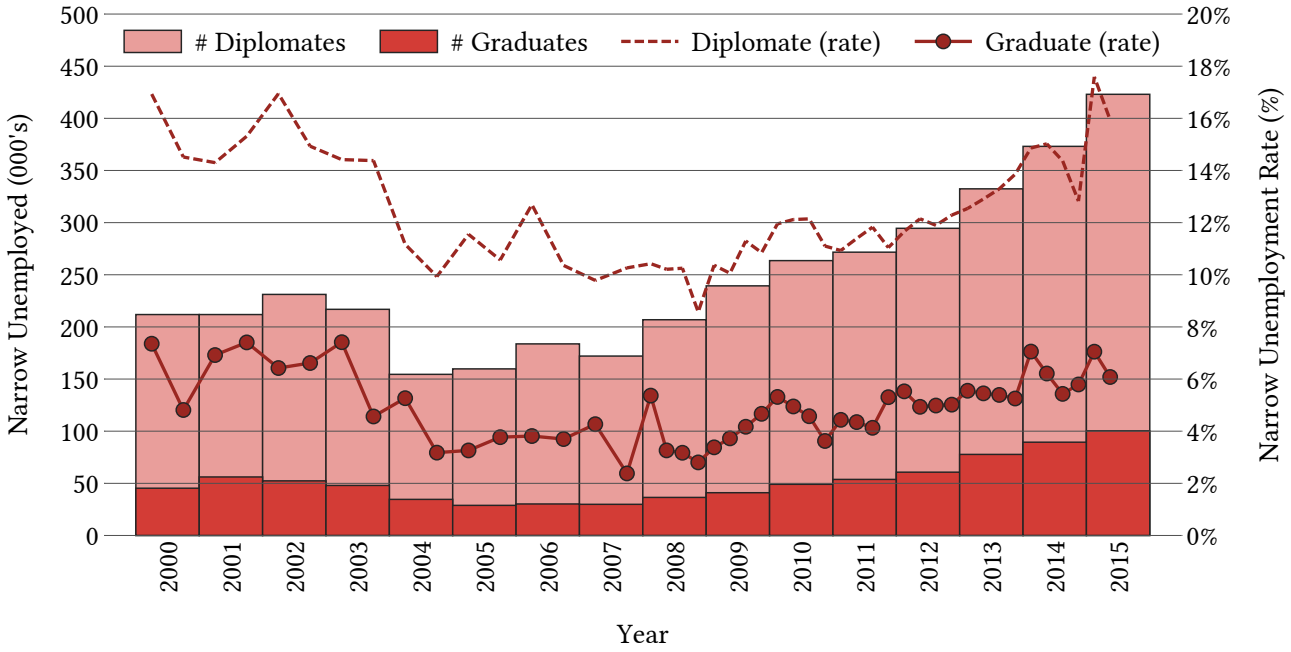
Figure 2: Employment and employment rates (%) for graduates and diplomates (2000 - 2015)



NOTES: Own estimations using Stats SA's March 2000 LFS - QLFS2015Q2 data. Estimates are weighted and are calculated only for graduates and dipolomates in the population of working-age (15 - 64 year-olds). Bars denote the respective numbers of employed graduates and diplomates and have been stacked (measured on the left-hand-side vertical axis). Lines denote the respective graduate and diplomate employment rates (measured on the right-hand-side vertical axis).

narrowly unemployed. Instead, the narrow unemployment rate for graduates has consistently been 5 percentage points or more lower, on average than the narrow unemployment rate for diplomates.

Figure 3: Narrow unemployment and narrow unemployment rates (%) for graduates and diplomates (2000 - 2015)



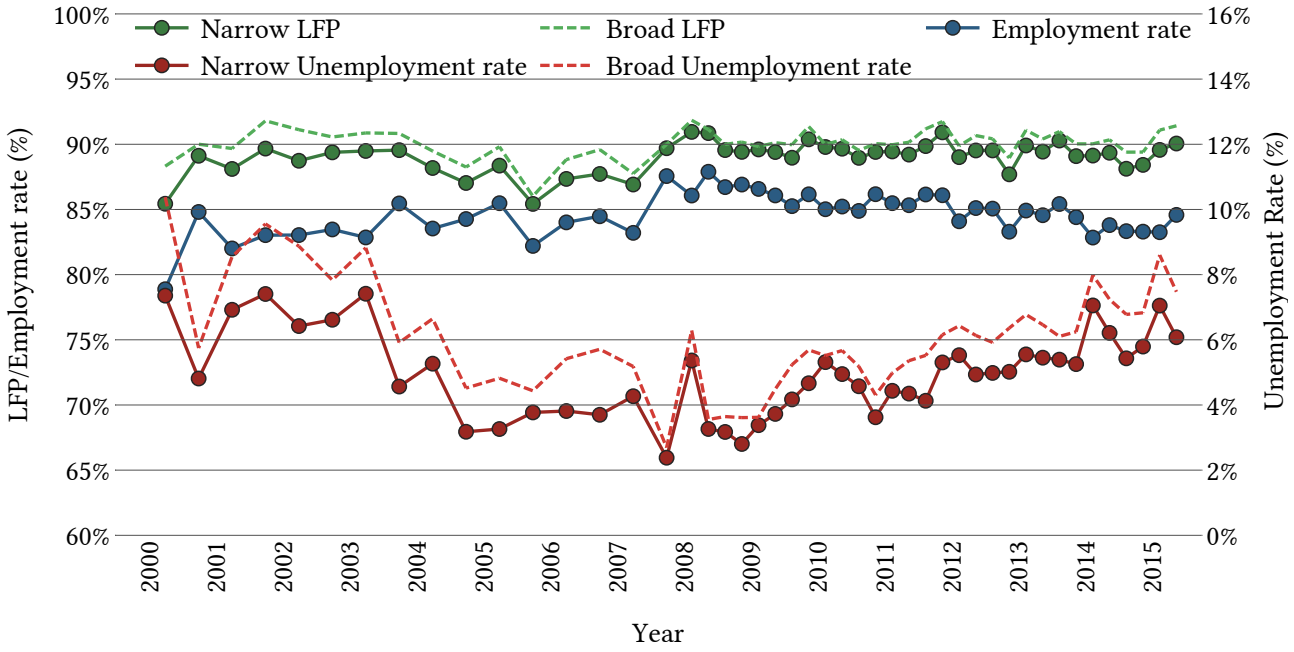
NOTES: Own estimations using Stats SA's March 2000 LFS - QLFS2015Q2 data. Estimates are weighted and are calculated only for graduates and dipolomates in the population of working-age (15 - 64 year-olds). Bars denote the respective numbers of narrowly unemployed graduates and diplomates and have been stacked (measured on the left-hand-side vertical axis). Lines denote the respective graduate and diplomate narrow unemployment rates (measured on the right-hand-side vertical axis).

These figures offer a simple, but compelling argument against the misguided practice of drawing inferences about the labour market outcomes faced by university graduates and degree-holders from the analysis of the labour force outcomes faced by tertiary-educated individuals as a whole. Doing so will clearly lead to upward-biased perceptions of graduate unemployment in the country. This is particularly poignant when one takes into account that most of the survey data on which analyses regarding employment and unemployment outcomes in South Africa are based do not enable one to distinguish between diplomates who obtained their qualifications from HEIs and diplomates who obtained their qualifications from TVET colleges.[7] Yet, it is not only known that TVET diplomates constitute a significant proportion of all diplomates, but also that the quality of TVET college diplomas and certificates are generally lower and, therefore, less likely to improve employment prospects and reduce the probability of unemployment, than HE diplomas and certificates (Financial and Fiscal Commission, 2012; Fisher and Scott, 2011). Therefore, there is a clear case to be made for analysing the labour market outcomes faced by university graduates separately from those faced by diplomates and to use the latter only as a comparator group.

Figure 4 presents the broad and narrow LFP, employment, and unemployment rates for graduates between 2000 and 2015. The distinction between the broad and narrow labour force often plagues studies concerned with unemployment (Kingdon and Knight, 2006). However, as can be seen from the graph, the difference between these two definitions of LFP and unemployment are mostly negligible for graduates. Therefore, it is largely irrelevant weather one analyses unemployment outcomes for graduates in South Africa using the

[7] In this paper, HE refers only to South Africa's public higher education system and thus excludes TVET colleges and private HEIs. Similarly, HEIs either refer to the 36 former technikons or general academic universities or the the 23 present-day universities that constituted South Africa's public HE system until 2014. It follows that *HE-educated individuals* refer only to those individuals who have completed either a diplomate or graduate-level qualification at one of South Africa's public HEIs.

Figure 4: Broad and narrow graduate LFP, employment, and unemployment rates (%) (2000 - 2015)



NOTES: Own estimations using Stats SA's March 2000 LFS - QLFS2015Q2 data. Estimates are weighted and are calculated only for graduates in the population of working-age (15 - 64 year-olds). LFP and employment rates are measured on the left-hand-side vertical axis, whereas unemployment rates are measured in the right-hand-side vertical axis.

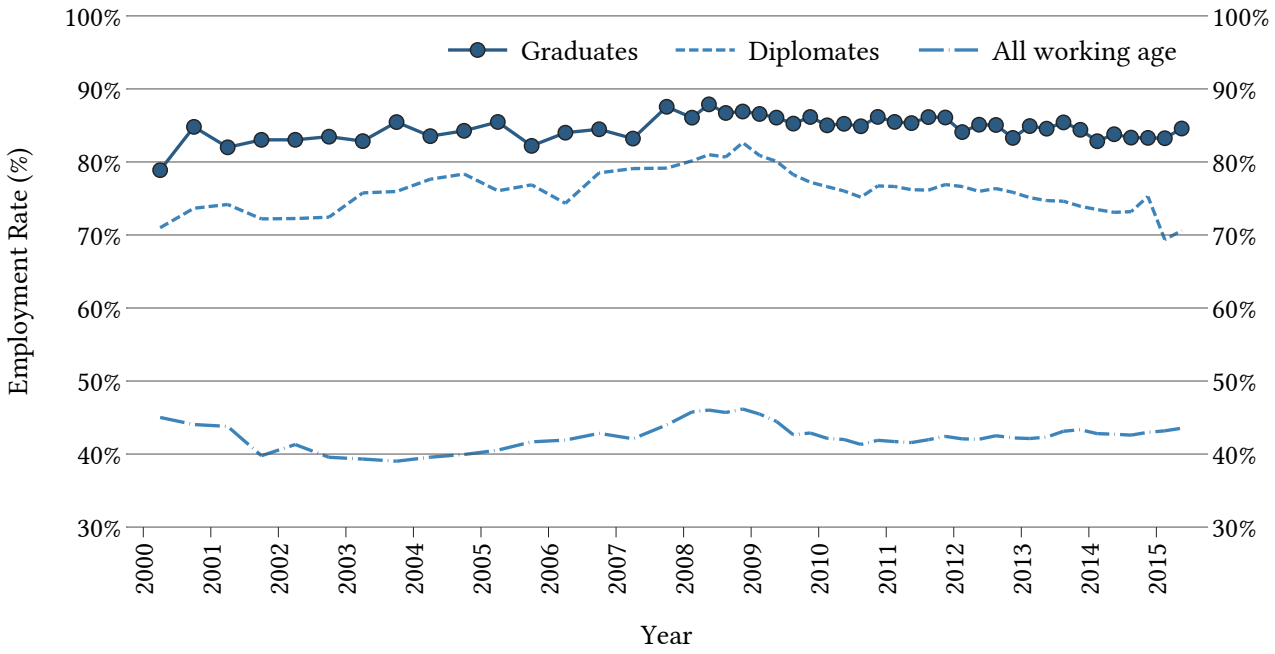
broad or the narrow definition of the labour force.

While graduate LFP rates have fluctuated slightly between 2000 and 2015, there is no clear evidence of a long-term upward or downward trend. By contrast, graduate employment rates appear to have risen between 2000 and 2008 and declined again thereafter with graduate unemployment rates largely mirroring this pattern. However, the trends in these labour market status outcomes for graduates is of secondary concern. Of primary importance is the fact that graduate employment rates have consistently been higher than 80% since 2001 and that, with the exception of 2014 and 2015, graduate unemployment rates have consistently been lower than 6% since September 2003.

The estimates in Figure 4 suggest that graduate unemployment in South Africa is not alarmingly high. In fact, it is rather low. Yet, in order to draw such a conclusion it is necessary to evaluate graduate employment and unemployment rates in the context of South Africa's overall employment and unemployment rates, as done in Figures 5 and 6. It is clear from the graphs that South Africa's overall employment rate in the population of working-age is extremely low at between 40% and 45%. Similarly, the overall narrow unemployment rate in the population of working age of around 25% is extremely high.

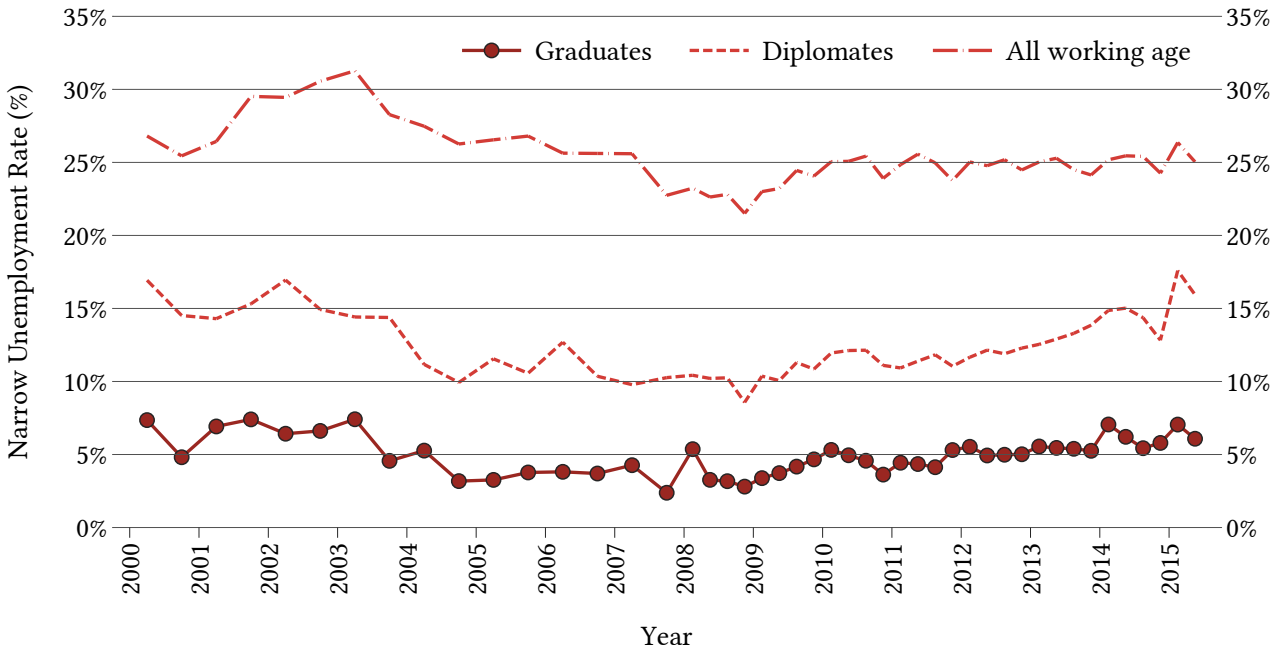
Within the context of South Africa's high overall unemployment rate and its low employment rate, it would be difficult to argue that graduate employment is low or that graduate unemployment is alarmingly high. Such assertions do not seem to have any basis in reality and are clearly not supported by the data. Crucially, however, overall figures fail to reflect the differences in graduate labour market status outcomes between race groups. Figure 7, for example, shows that, while the narrow LFP and employment rates for White graduates track together very closely over time, there has been a far larger difference of around 5 percentage points between the narrow LFP rate and the employment rate for Black graduates since 2004.

Figure 5: Employment rates (%) for graduates, diplomates, and the population of working-age (2000 - 2015)



NOTES: Own estimations using Stats SA's March 2000 LFS - QLFS2015Q2 data. Estimates are weighted and are calculated only for the population of working-age (15 - 64 year-olds).

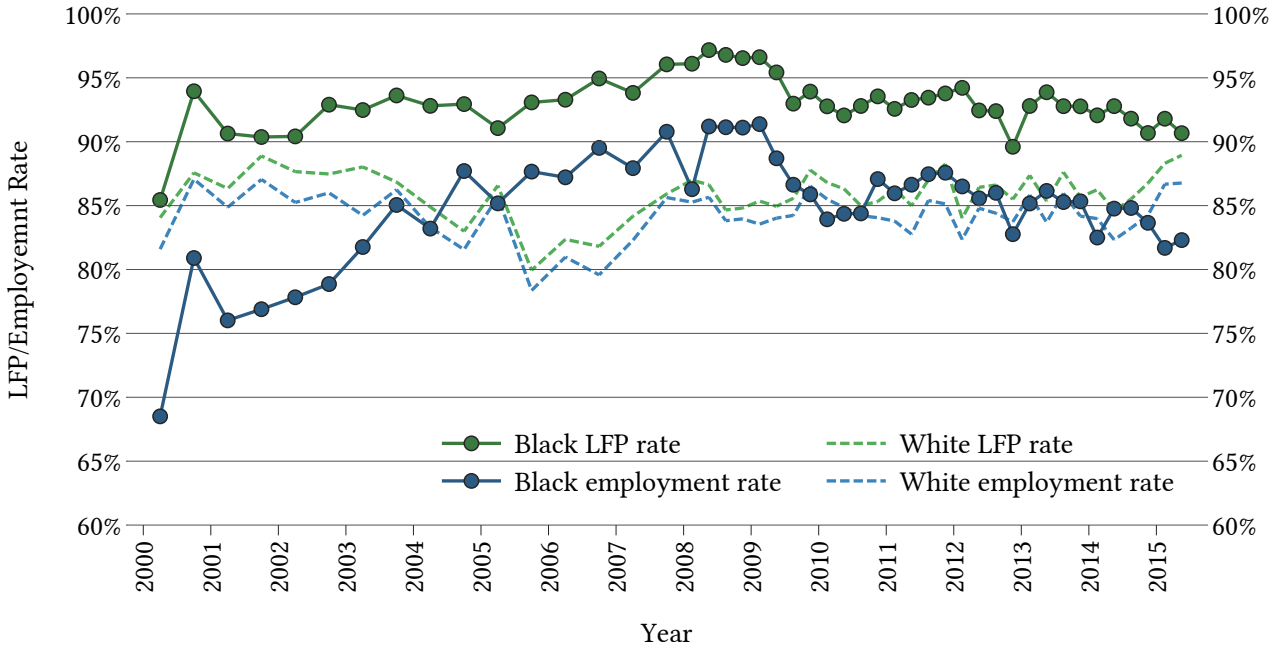
Figure 6: Narrow unemployment rates (%) for graduates, diplomates, and the population of working-age (2000 - 2015)



NOTES: Own estimations using Stats SA's March 2000 LFS - QLFS2015Q4 data. Estimates are weighted and are calculated only for the population of working-age (15 - 64 year-olds).

Despite the fact that the employment rate for Black graduates has been at least as high as the employment rate for White graduates since 2003, the fact that Black graduates have a significantly higher narrow LFP rate than White graduates means that they also have a significantly higher narrow unemployment rate. This is illustrated in Figure 8. While the narrow unemployment rate for Black graduates has decreased considerably

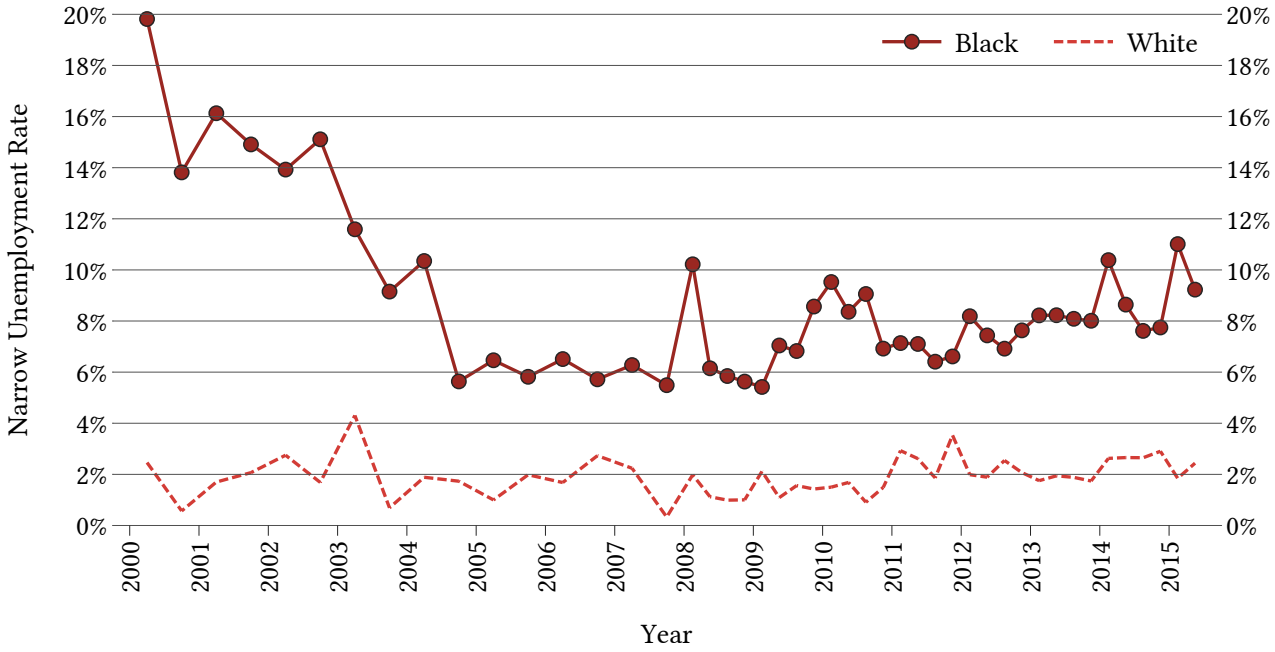
Figure 7: Narrow LFP and employment rates (%) for Black and White graduates (2000 - 2015)



NOTES: Own estimations using Stats SA's March 2000 LFS - QLFS2015Q2 data. Estimates are weighted and are calculated only for the population of working-age (15 - 64 year-olds). Lines denote the estimated narrow LFP and employment rates for Black and White graduates, respectively.

over time, from an estimated high of 20% in 2000 to around under 10% in 2015, it nevertheless remains more than twice as high as the unemployment rate for White graduates.

Figure 8: Narrow unemployment rates (%) for Black and White graduates (2000 - 2015)



NOTES: Own estimations using Stats SA's March 2000 LFS - QLFS2015Q2 data. Estimates are weighted and are calculated only for the population of working-age (15 - 64 year-olds). Lines denote the estimated narrow unemployment rates for Black and White graduates, respectively.

None of the findings in this section can be taken to suggest that graduate unemployment in South Africa is

either alarmingly high, or that it is rising at an alarming rate. On the contrary, it appears as though graduate unemployment rates have been on a long-term downward trend since 2000 and are low in relation to overall unemployment in the country. Moreover, the gap between the unemployment rate for Black and White graduates has narrowed considerably between 2000 and 2005. Yet, the fact that such a gap still exists and that it does not appear to be narrowing after 2005 begs the question: what it is that distinguishes Black and White graduates, such that the former group is likely to face worse labour market outcomes than the latter group? More generally, it remains unclear why there are unexplained differences in the employment and unemployment outcomes for graduates from different race groups and how these differences relate to HE institutional factors. The objective in the remainder of this paper is to provide answers to this question.

4 The South African HE landscape

Historically, South Africa's HE landscape has been highly fragmented. Prior to 2004, the public HE system consisted of 36 HEIs, comprising 15 technikons and 21 general academic universities. While technikons operated as *de facto* vocational training institutions, focussing primarily on the application of knowledge, universities concentrated on the development of knowledge and the training of students in such scientific and scholarly disciplines as would enable them to occupy high-level professions (Bunting, 2002:37 -39). However, the HE system was not only fragmented in terms of function, but also in terms of governance, funding and, as a result, the quality of education provided by different parts of the system (CHE, 2004:24).

Under Apartheid, eight racially demarcated government departments were tasked with the administration of the 36 HEIs. Significant differences in the amount of funding and resources available to each department and the amount of developmental support they were therefore able to provide the various HEIs under their control, meant that this policy had the effect of further fragmenting the HE system into what can most accurately be described as historically disadvantaged, or Black, institutions (HDIs) and historically advantaged, or White, institutions (HAIs) (CHE, 2004:xv). The classification of each of South Africa's 36 former HEIs as either historically disadvantaged or historically advantaged is shown in Table A.2. In total, 10 of the former universities and 7 of the former technikons can be regarded as historically disadvantaged.

Following South Africa's democratization in 1994, the HE landscape was subjected to a number of significant policy changes, chief among which was the amalgamation of its 36 technikons and universities into 11 traditional universities, 6 comprehensive universities, and 6 universities of technology (CHE, 2010:2).[8] This amalgamation not only reduced the total number of public HEIs from 36 to the current 23 HEIs shown in Table A.3, but also meant that some technikons merged with general academic universities and, more importantly, that some HDIs merged with HAIs.[9]

Despite the aforementioned policy changes, it is important to acknowledge that the South African HE system remains fragmented along the lines of historical advantage and disadvantage. Many HDIs are still at a significant disadvantage relative to HAIs in terms of their institutional capacities, the socio-economic backgrounds

[8] Traditional universities and universities of technology respectively resemble the pre-amalgamation general academic universities and technikons in function, with the former offering mainly theoretically-oriented diplomas and degrees and the latter mainly vocational diplomas and degrees. Comprehensive universities offer a combination of these types of qualifications .

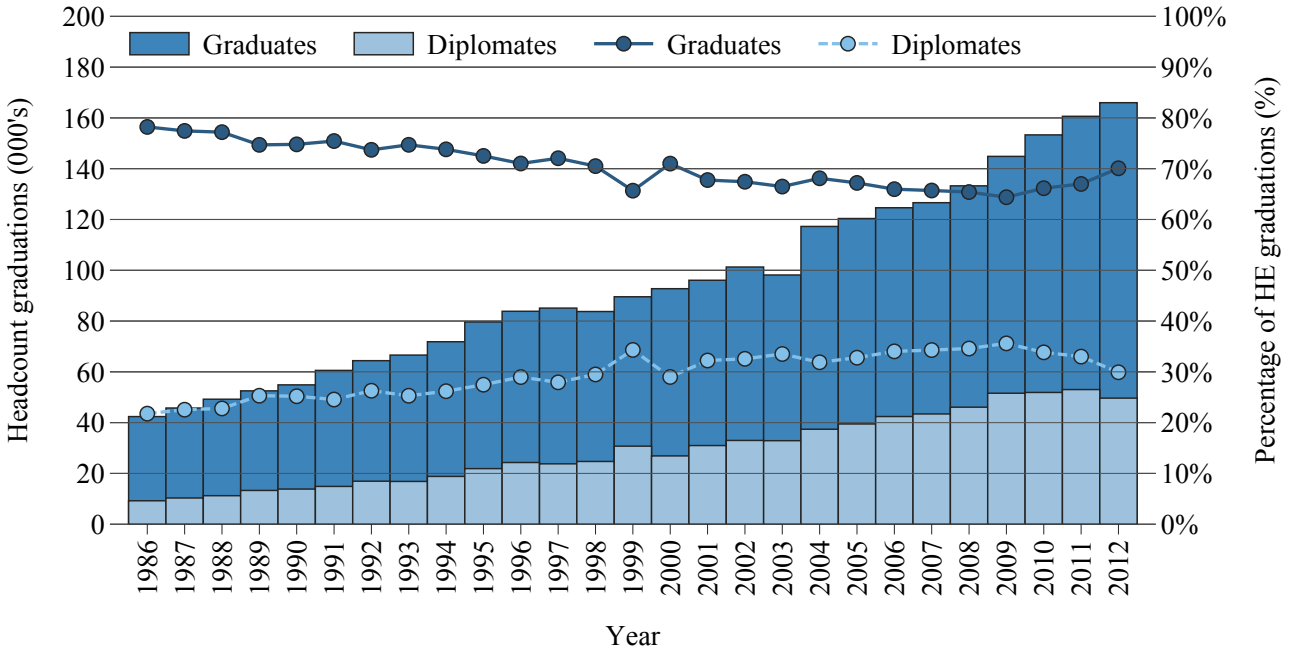
[9] The present-day Cape Peninsula of Technology (CPUT), Durban Institute of Technology (DUT), University of Kwazulu-Natal (UKZN), North West University (NWU), and Tshwane University of Technology (TUT), for example, were all created from the merger of historically disadvantaged and historically advantaged HEIs.

of their students, and the quality of education that they can provide. Consequently, it is reasonable to expect that HDIs and HAIs will perform differently, not only in terms of graduate outputs, but also in terms of many other performance measures (Fisher and Scott, 2011:28).

4.1 Changes in HE graduate outputs (1986 - 2012)

The policy changes which have altered the South African HE landscape over the three decades have coincided with a significant rise in the total number of HE graduations each year. Figure 9 shows that, while only just over 40 000 individuals graduated from HEIs with university or technikon qualifications in 1986, this number had more than doubled by 1996. Following a period of relative stagnation between 1996 and 2000, the number of diplomates and graduates produced annually again began to rise rapidly and by 2012 South Africa's 23 universities produced just over 165 000 HE-educated individuals each year. However, as can also be seen from Figure 9, the number of HE-educated individuals with diplomate-level qualifications has been rising faster than the number of individuals with graduate-level qualifications. Where approximately 3.5 graduates were produced for each diplomate in 1986, this ratio had fallen to just over 2.3 graduates per diplomate by 2012. Thus, while graduates still represent the bulk of HE-educated individuals produced by universities each year, their relative share of South Africa's stock of HE-educated individuals is steadily declining.

Figure 9: Graduate- and diplomate-level graduations (1986 - 2012)

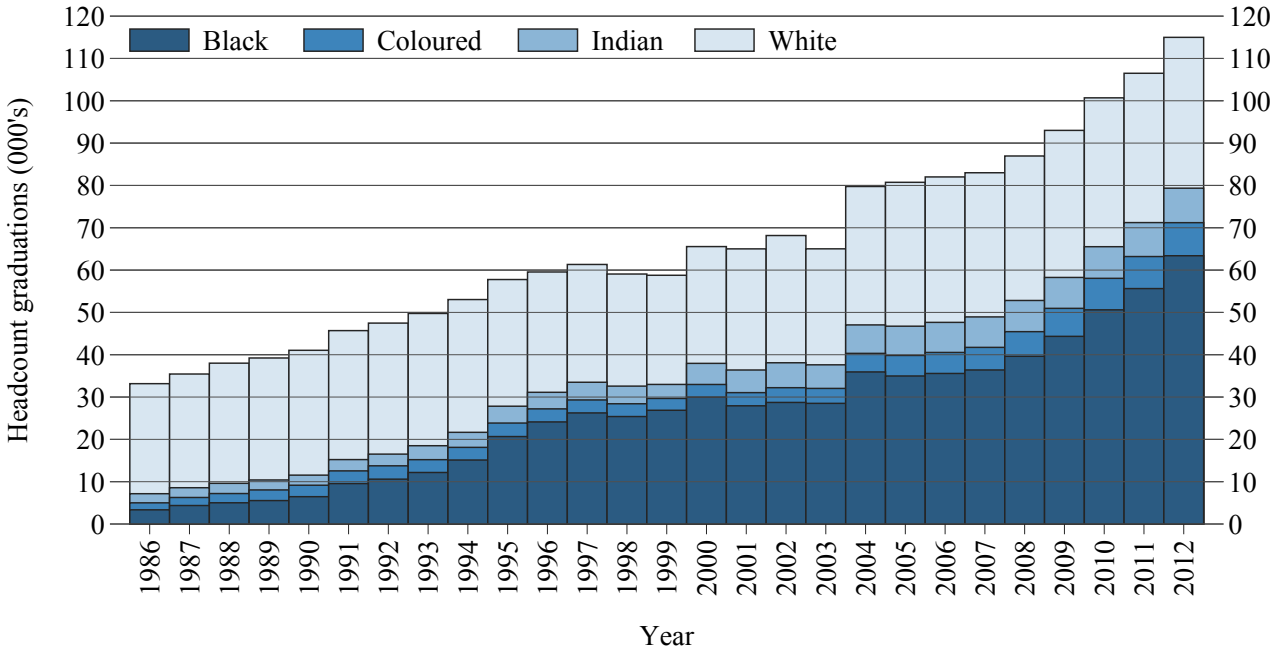


NOTES: Own calculations using aggregate HEMIS data accessed via DHET (2014). Bars denote the respective numbers of diplomate- and graduate-level graduations per year and have been stacked (measured on the left-hand-side vertical axis). Lines denote the respective diplomate and graduate shares of all HE graduations in the public HE system per year (measured on the right-hand-side vertical axis)

In addition to the expansion of South Africa's yearly graduate outputs, the nature of the policy changes which have affected the HE system over the past 25 years means that the demographic composition of South Africa's stock of graduates has also changed radically over time. This is clearly evident when looking at changes in the racial composition of the graduates produced by the HE system each year. Figure 10 reveals that, while the number of White graduates produced annually has increased only moderately from about 27 500 to just over

35 000 in the past 25 years, the number of Black graduates produced has increased more than 16-fold from about 3 400 in 1986 to more than 63 000 in 2012. The implications of the racial differences in graduate output growth are simple: while the HE system produced 7.9 White graduates for each Black graduate in 1986, by 2012 it produced 1.8 Black graduates for every single White graduate. Figure 11 offers a similarly poignant illustration of the extent of change in the racial composition of South Africa's stock of graduates by showing the respective racial shares of the total number of graduates produced in each year since 1986.

Figure 10: Graduate-level graduations, by race (1986 - 2012)



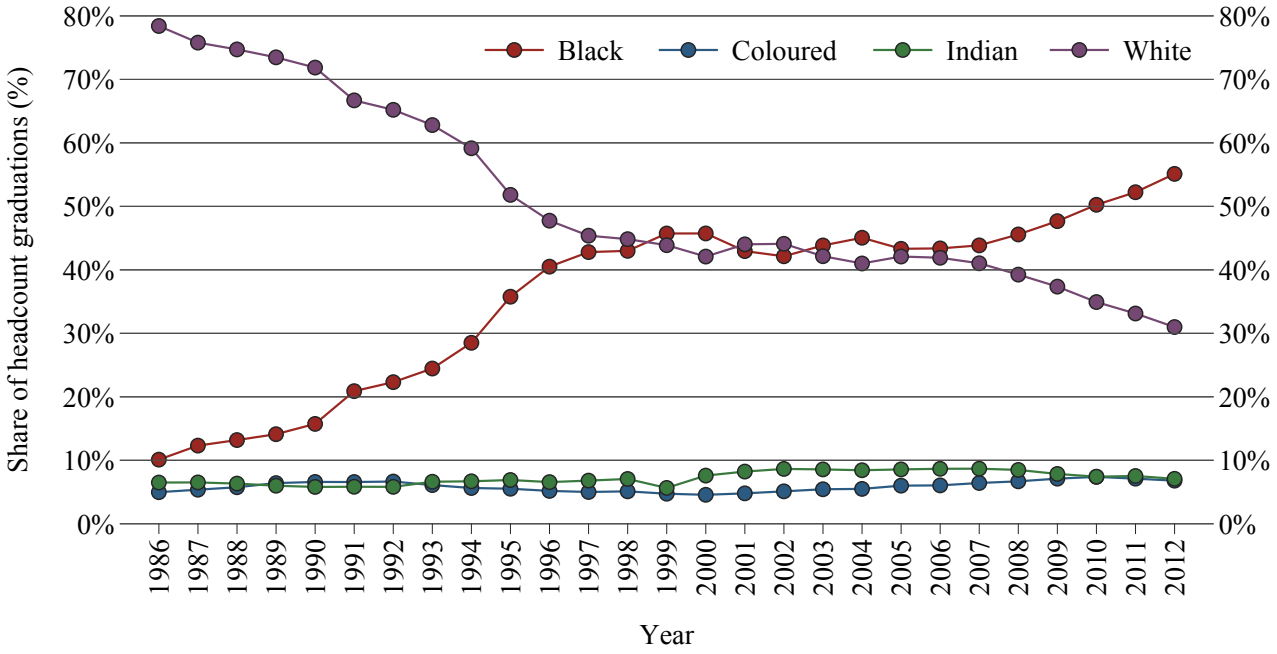
NOTES: Own calculations using aggregate HEMIS data accessed via DHET (2014). Bars denote the respective numbers of graduate-level graduations in the public HE system per year for Black, Coloured, Indian, and White individuals and have been stacked.

Figure 12 shows how the amalgamation of technikons and universities in South Africa in 2004 impacted on the relative contributions made by different types of HEIs to total annual graduate outputs. Prior to 2004, universities accounted for around 90% of all graduate-level graduations each year. However, since 2004, only about 60% of all graduations have come from traditional universities, with 30% now being produced by comprehensive universities. Given that all universities of technology either used to be technikons or were created through the merger of technikons, it is not surprising that this part of the HE system still contributes only about 10% of graduate-level graduations every year, just as it did before the amalgamation.

As mentioned before, the amalgamation of South Africa's 36 former HEIs not only had the effect of reducing the total number of HEIs in the country, but also entailed that some HDIs merged with HAIs. From the perspective of analysing the relative contributions of the historically disadvantaged and historically advantaged parts of the HE system to the total number of graduates produced each year, this is problematic since it is no longer clear to what extent these institutions can accurately be classified as either HDIs or HAIs. This problem is illustrated in Figure 13, which shows the respective HDI and HAI shares of graduate-level graduations.

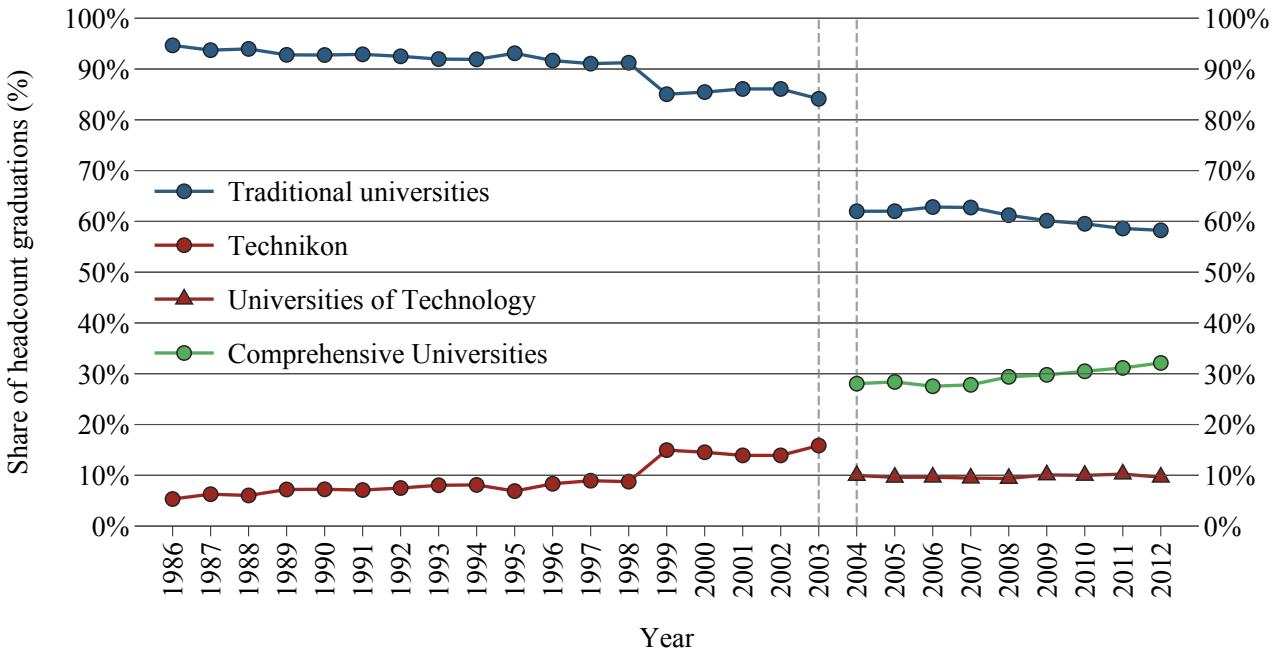
HAIs and HDIs respectively produced around 80% and 20% of South Africa's graduate-level graduations by 2003. However, if one applies the classification commonly used in the literature on South Africa's HE system, whereby institutions that were either already classified as historically disadvantaged before 2004 or were

Figure 11: Racial shares of graduate-level graduations (1986 - 2012)



NOTES: Own calculations using aggregate HEMIS data accessed via DHET (2014). Lines denote the respective racial shares of all graduate-level graduations in the public HE system per year.

Figure 12: Graduate-level graduation shares, by HEI type (1986 - 2012)



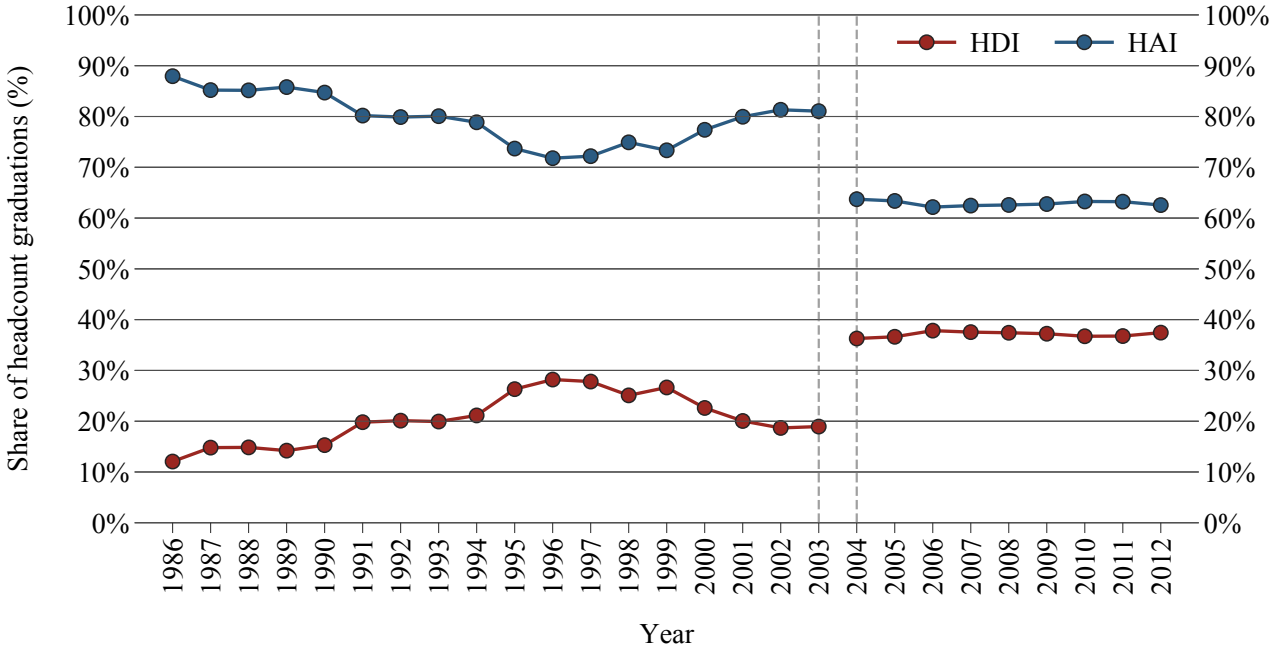
NOTES: Own calculations using aggregate HEMIS data accessed via DHET (2014). Lines denote the respective shares of all graduate-level graduations per year in the public HE system for technikons, traditional universities, universities of technology, and comprehensive universities.

merged with HDIs as part of the amalgamation are now also be described as HDIs, there is a large, discontinuous change in the relative contributions of HDIs and HAIs.[10] Specifically, this classification makes it seem

[10] According to this classification, the new HDIs include 12 institutions: University of Fort Hare (UFH), University of KwaZulu-Natal (UKZN), University of Limpopo (UL), North West University (NWU), University of Venda (UNIVEN), University of Western Cape (UWC), University of Zululand (UZ), Walter Sisulu University (WSU), Cape Peninsula University of Technology (CPUT), Durban

as though HDIs have been producing just short of 40% of all new graduates since 2004.

Figure 13: Graduate-level graduation shares, by HDIs vs HAIs (1986 - 2012)



NOTES: Own calculations using aggregate HEMIS data accessed via DHET (2014). Lines denote the respective shares of all graduate-level graduations per year in the public HE system for HAIs.

Due to the potential pitfalls inherent in using a classification which is based solely on historical status to evaluate post-amalgamation HEIs , CHET (2010) proposes a three-cluster classification of South Africa's universities which expresses institutional differentiation in terms of observable criteria and performance measures (Fisher and Scott, 2011:33).[11] As shown in Table A.3, the first cluster comprises South Africa's leading research institutions, all of which are HAIs. Cluster 2 is composed of both traditional and comprehensive universities while the third cluster includes all the universities of technology, most of which could be classified as HDIs, and two comprehensive universities (Fisher and Scott, 2011:33). Though the original aim of the 3-cluster classification was to differentiate HEIs based on function and focus, it nevertheless provides a useful hierarchical classification of institutional quality in different parts of the HE system.

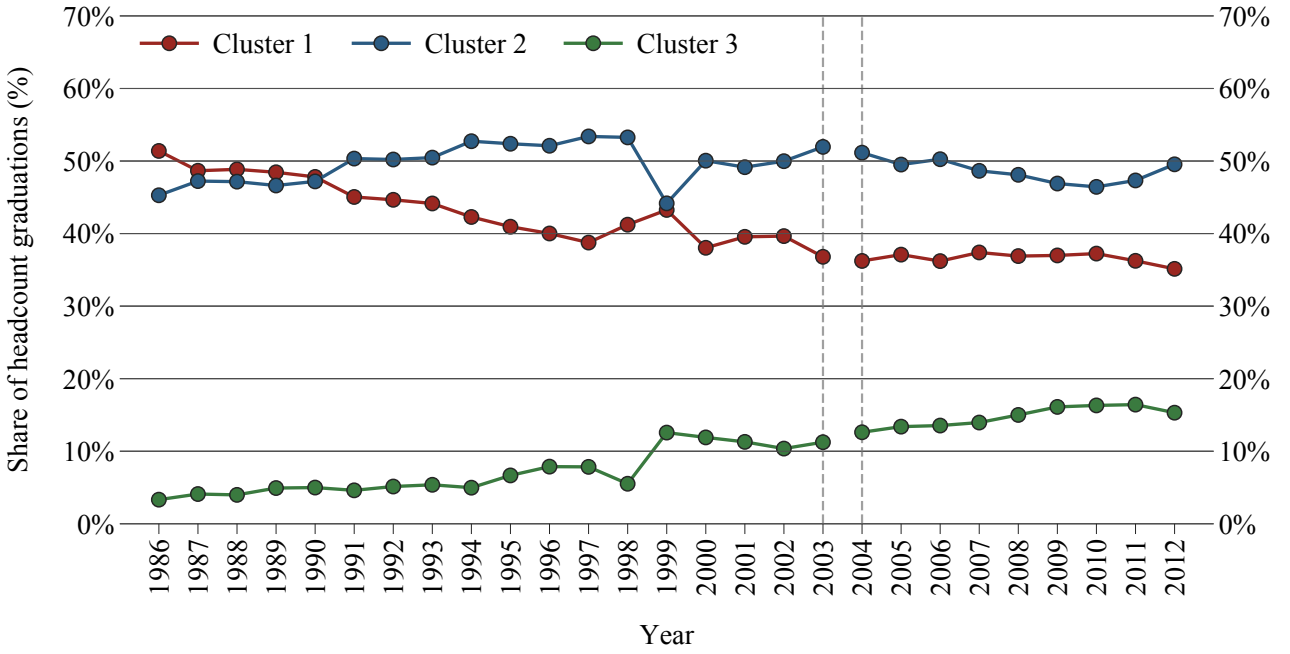
Figure 14 shows the shares of total graduates produced each year by universities in the three different HEI clusters.[12] In the long run, cluster 3 institutions have been increasing their graduate outputs relative to cluster 1 universities. In the last 10 years, however, cluster 3 institutions have been increasing their graduate outputs relative to both cluster 1 and cluster 2 institutions. By 2015, 50% of new graduates were being produced by cluster 2 universities, followed by 35% by cluster 1 universities and 15% by cluster 3 universities.

Institute of Technology (DUT), Tshwane University of Technology (TUT), and Mangosuthu University of Technology (MUT).

[11] The observable input criteria used in the construction of the three CHET (2010) HE institutional clusters include: the percentage headcount enrolment in science, engineering and technology; the percentage master and doctoral headcount enrolments; the student to academic and/or research staff FTE ratio; the percentage of permanent academic and/or research staff with doctoral degrees; the percentage private income; and the government and/or student fee income per FTE student. The performance measures used in the construction of the clusters include student success rates, graduation rates, and the weighted research outputs units per permanent academic and research staff member.

[12] The CHET (2010) cluster classification was retrospectively applied to the 36 pre-amalgamation technikons and universities based on the HEIs into which they were merged in 2004.

Figure 14: Share of annual graduate-level graduations, by HEI cluster (1986 - 2012)



NOTES: Own calculations using aggregate HEMIS data accessed via DHET (2014). Lines denote the respective shares of all graduate-level graduations per year in the public HE system for cluster 1, cluster 2, and cluster 3 HEIs (CHET, 2010).

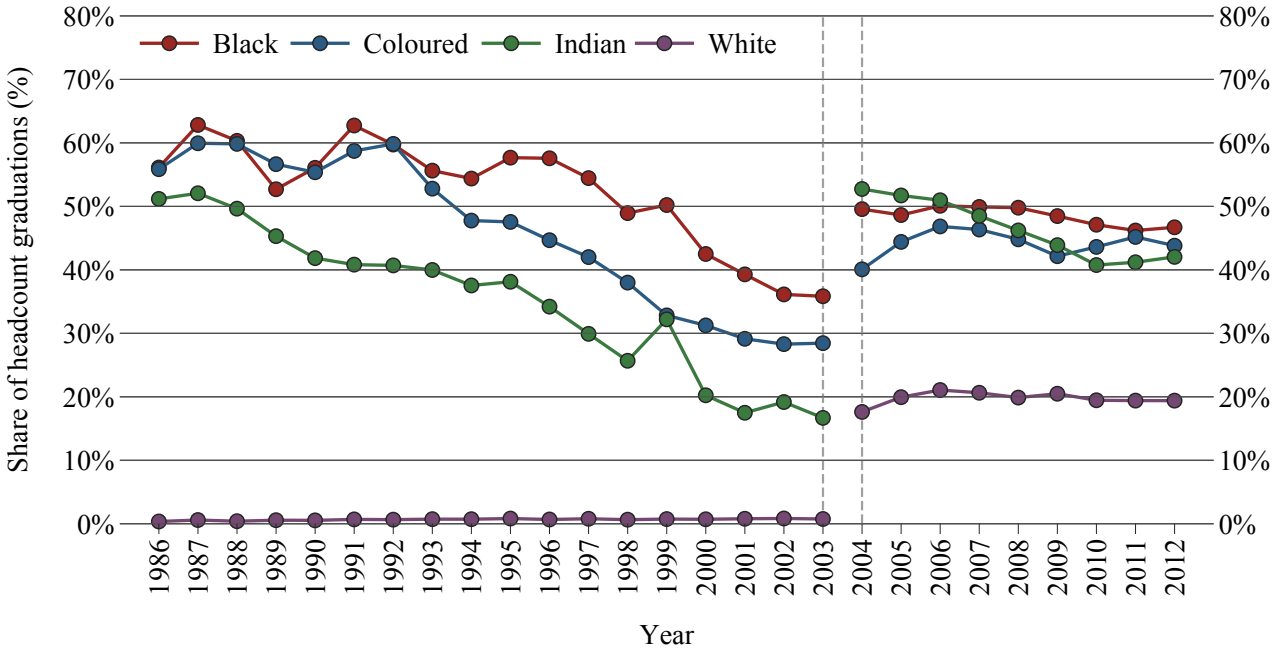
The racial dimensions of historical status in South Africa's HE system coupled with the significant expansion of the number of Black graduates produced by the country's HEIs over the past 25 years imply that the aforementioned changes in the HE landscape are unlikely to have been equally pertinent to all race groups. This is confirmed by Figure 15 which shows marked differences in the proportions, and changes in the proportions, of Black, Coloured, Indian, and White graduates produced by HDIs.

In 1986, more than 50% of Indian, Coloured, and Black graduates graduated from HDIs. By 2003, the percentage of Black graduates from HDIs had fallen to 35%, the percentage of Coloured graduates from HDIs to 29%, and the proportion of Indians from HDIs to 18%. Crucially, this change was not driven by a decline in the numbers of Black, Coloured, and Indian graduates being produced by HDIs. Rather, it was the result of the fact that the number of Black, Coloured, and Indian students who graduated from HAIs increased comparatively more rapidly between 1986 and 2003. Ignoring what is most likely a definition-driven discrete jump in the proportion of graduates from HDIs across all race groups between 2003 and 2004, it appears as though the historical downward trend in the proportion of Black and Indian graduates from HDIs has continued in the years following the amalgamation.

Despite the general decline in the HDI-share of graduations, Figure 15 suggests that a far greater proportion of Black, Coloured, and Indian graduates still graduate from historically disadvantaged HEIs than is the case for Whites. This supposition is supported by Figure 16 which shows that, while 57% and 39% of White graduates respectively graduated from cluster 1 and cluster 2 HEIs in 2015, a mere 5% graduated from cluster 3 HEIs. By contrast, in the same year over 56% of Black graduates graduated from cluster 2 institutions and the percentage of Black graduations from cluster 1 or cluster 2 HEIs was roughly equal at about 22% each.

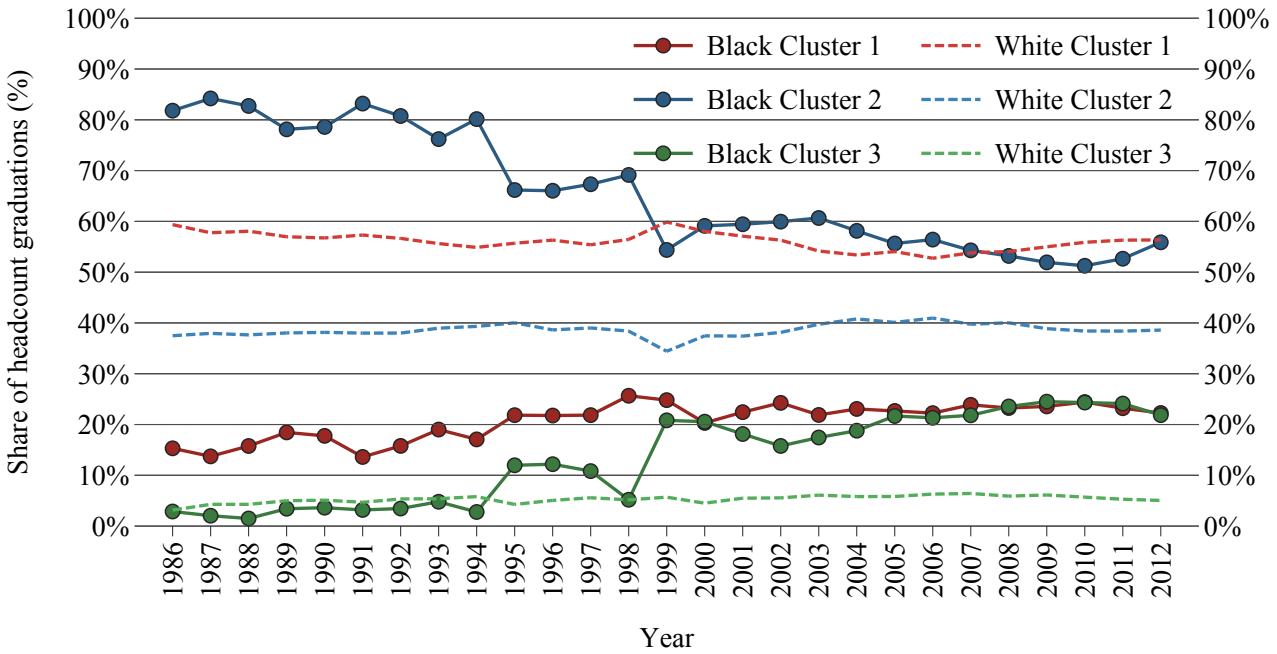
It is reasonable to expect that the various features of South Africa's HE system and the changes in the HE landscape outlined above would have important implications for the labour market prospects faced by the

Figure 15: HDI-share of Black, Coloured, Indian, and White graduate-level graduations (1986 - 2012)



NOTES: Own calculations using aggregate HEMIS data accessed via DHET (2014). Lines denote the respective HDI shares of all Black, Coloured, Indian, and White graduate-level graduations per year in the public HE system.

Figure 16: Share of Black and White graduate-level graduations by HEI cluster (1986 - 2012)



NOTES: Own calculations using aggregate HEMIS data accessed via DHET (2014). Lines denote the respective shares of all Black and White graduate-level graduations per year in the public HE system for cluster 1, cluster 2, and cluster 3 HEIs (CHET, 2010).

country's graduates. In the absence of a commensurate increase in the demand for graduate labour and expansion of the labour market's capacity to absorb graduates into graduate-level jobs over the past 25 years, the rapid rise in the number of graduates produced by the HE system each year should mean that new graduates find it increasingly difficult to procure employment. Second, and perhaps more importantly, the significant

expansion of South Africa's stock of Black graduates, in particular, must be viewed in the context of historically limited access to quality HE. That is, because of historical inequalities in access to quality education, the fact that the Black share of graduate-level graduations is rising over time also means that South Africa's stock of graduates are increasingly being supplemented by individuals who are likely to have been educated in the weaker-performing parts of the HE system. Thus, it is plausible that part of the difference in the unemployment rates that are observed for Black and White graduates could be attributed to the fact that a far greater proportion of Black graduates (more than 78%) graduate from cluster 2 or 3 HEIs, for example, than White graduates, the majority of which graduate from cluster 1 institutions.

5 Relating HEIs to graduate unemployment and employment probabilities

As discussed above, historical patterns of access to HEIs, persistent heterogeneity in the type and quality of university education, and the changing demographic composition of the country's stock of graduates are likely to be important for explaining racial differentials in graduate labour market outcomes in South Africa. However, no study has thus far been able to examine on a nationally representative basis the extent to which the nature of the specific HEIs attended by graduates is associated with the probabilities that they will be employed or unemployed. This is largely attributable to the fact that there is no existing dataset for South Africa that allows information on the HEIs attended by graduates to be linked directly to the labour market outcomes they face.[13] By implication, the success of any attempt to empirically investigate the relationship between HE institutional features and graduate labour market outcomes in South Africa hinges on the extent to which it is possible to link" information regarding graduate labour market outcomes in one dataset, to information regarding graduate HE institutional aspects in another dataset.

5.1 Data

The analysis below exploits two distinct sources of data on South African graduates. The first is a pooled sample of cross-sectional labour force data for working-age graduates from Statistics South Africa's (StatsSA) March and September 2000- 2007 Labour Force Surveys (LFS) and its 2008Q1 - 2015Q2 Quarterly Labour Force Surveys (QLFS). The second source of data comes from the Department of Higher Education and Training's (DHET) Higher Education Management Information System (HEMIS).

HEMIS is the national repository for information on students who have enrolled in and subsequently graduated from the public HE system in South Africa and, in its original form, contains detailed unit-record information on all enrolments and graduations since 2000. The HEMIS data used in this paper, however, while based on the aforementioned unit-record information, has been aggregated in such a way that it is no longer possible to identify individual student records.[14] Nevertheless, the data contains sufficiently detailed information on student demographics and the specific HEIs where different graduates obtained their qualifications to be used for the purposes of the empirical methodology outlined below.

[13] On the one hand, none of the nationally representative labour force survey datasets available for South Africa contain information on the tertiary institutions where graduates obtained their qualifications and, on the other hand, HE administrative records containing detailed information on the individuals who have graduated from public HEIs in South Africa do not contain any information on the labour market outcomes subsequently faced by those graduates.

[14] This aggregate HEMIS data was extracted from the IDSC's Higher Education Data Analyser (HEDA, 2015).

While the pooled labour force survey data (hereafter collectively referred to as *LFS data*) covers the period 2000 - 2015, audited aggregate HEMIS data is currently only available for the period 2000 - 2013.

5.2 Methodology

In order to examine the association between HEIs and graduate employment and unemployment probabilities, it is first necessary to find a way of linking the information on graduates in the HEMIS data to information on graduates in the LFS data.

The approach proposed here combines forms of multiple imputation and probabilistic cell-matching and entails using the availability of common time-invariant group-specific variables found in both the LFS and HEMIS data to estimate the probability that specific LFS graduates come from specific groups of HEMIS graduates.[15] Specifically, by using information that is unique across different combinations of time-invariant group-specific variables in both the HEMIS and LFS data, the approach exploits the fact that it is theoretically possible to assign to each graduate in the LFS data an estimated probability of having graduated from a specific South African HEI, based on the known distribution of graduations in the HEMIS data.

The time-invariant group-specific variables that are common across the HEMIS and LFS data can be represented by a series of vectors, $\mathbf{X}^H, \mathbf{Y}^H \ldots \mathbf{Z}^H$ and $\mathbf{X}^L, \mathbf{Y}^L \ldots \mathbf{Z}^L$, where the superscripts H and L respectively denote the HEMIS and LFS datasets. Consequently, $\mathbf{x}_i^j$ would denote the i^{th} observation of variable $\mathbf{X}$ in dataset j.

There is a finite number of unique combinations of observed values that the set of group-specific variables takes on in each dataset. It is therefore possible to construct a criterion index variable, $\mathbf{c}$, that uniquely identifies each of the unique combinations that occurs in either dataset. That is, $\mathbf{c}_i^j \in \mathcal{C}$ where $\mathcal{C}$ is the set of indices of unique patterns in $\left\{\mathbf{X}^H, \mathbf{Y}^H, \ldots, \mathbf{Z}^H\right\} \bigcup \left\{\mathbf{X}^L, \mathbf{Y}^L, \ldots, \mathbf{Z}^L\right\}$. In other words:

$$\mathbf{c}_i^j = \mathbf{c}_m^k \text{ if and only if } \left(\mathbf{x}_i^j = \mathbf{x}_m^k \text{ and } \mathbf{y}_i^j = \mathbf{y}_m^k \text{ and } \ldots \text{ and } \mathbf{z}_i^j = \mathbf{z}_m^k\right)$$

$$1^m\left(X^L = x_m^L, y^L = y_m^L, \ldots, Z^L = z_m^L\right) = c_m^L \quad \forall \quad m \in M$$

Let HEI be an index that takes on values in a set U that identifies the HEI from which individual i in the HEMIS data graduated. Calculate for each unique value of the index $\mathbf{c}$ in the HEMIS data, the proportion of graduates who graduated from a specific HEI, u. Call this variable $\mathbf{p_u}^j$ in dataset j.

$$\mathbf{p_u}_i^H = \Pr\left(HEI_i = u | \mathbf{c}_i^H = c\right) = \frac{\sum_{j=1}^N \mathbf{1}\left(\mathbf{c}_j^H = c\right)\mathbf{1}\left(HEI_j = u\right)}{\sum_{k=1}^N 1\left(\mathbf{c}_k^H = c\right)} \quad (1)$$
$$\forall\ (u, c) \in U \times \mathcal{C}^H$$

where $\mathbf{1}\left(.\right)$ denotes an indicator function.[16]

Wherever the index of unique patterns matches between datasets, assign to that observation in the LFS data the $\mathbf{p_u}$ value in the HEMIS dataset constructed as per equation (1). If a particular pattern in the LFS data does

[15] The methodology proposed here is based on the approaches discussed in Ridder and Moffitt (2007), Kim and Chambers (2012), Hof and Zwinderman (2012), and Goldstein *et al.* (2012).

[16] Note that there is one variable for each HEI represented in the HEMIS dataset. This effectively entails averaging the variables of interest (i.e. the specific university attended) over each unique value of the criterion.

not have a counterpart in the HEMIS data, a missing value is recorded.

$$\mathbf{p_u}_j^L = \begin{cases} \mathbf{p_u}_i^H & \text{if } \mathbf{c}_j^L = \mathbf{c}_i^H \\ \oslash & \text{otherwise} \end{cases}$$

For the sake of brevity, this approach is hereafter referred to as *probabilistic linking* while the imputed HEI probability variables, $\mathbf{p_u}_j^L$ are referred to as HEI proxies.

It should be clear that the accuracy of the probabilistic linking approach depends on the extent to which the values taken by the criterion, $\mathbf{c}$, uniquely identify the different observations in the LFS and the HEMIS data and sufficiently discriminate between graduates who graduated from different HEIs (Goldstein *et al.*, 2012:3481). This, in turn, is a function of the number of unique possible combinations of the identifier variables in relation to the total number of observations in each sample under consideration as well as the amount of variation in the number of distinct HEIs within each combination of the identifier variables.

Due to the fact that the questions regarding the highest education qualifications held by respondents in the LFS and QLFS changed between 2000 - 2015, three nested criteria had to be used sequentially to probabilistically link LFS graduates to HEMIS HEIs.[17] *Criterion 1* - the strictest criteria - consisted of unique combinations of respondents'/students' year of birth, race, gender, the type of graduate qualification held or awarded (e.g. a bachelors degree, post-graduate diploma, or master's degree or higher qualification), and the broad field of study in which the highest qualification was attained.[18] As no *field of study* questions were asked in the 2008Q1 - 2012Q2 QLFSs, *criterion 2* consisted of unique combinations of respondents'/students' year of birth, race, gender, and the type of graduate qualification held or awarded (e.g. a bachelors degree, post-graduate diploma, or master's degree or higher qualification). Finally, *criterion 3* consisted only of unique combinations of respondents'/students' year of birth, race and gender. In al cases, an attempt was made to first link on *criterion 1*, then on *criterion 2* and, in the event that a link still had not been established, on *criterion 3*.

Given that the HEMIS data used in this paper was only available for the period 2000 - 2013 and that it is not known when graduates observed in the LFS data graduated from the HEIs they attended, the probabilistic linking approach implicitly assumes that all LFS graduates for the period 2000 to 2015 were drawn from the 2000 to 2013 HEMIS graduation probability distribution. Put differently, the approach assumes that the conditional probability of having graduated from a specific HEI before 2000 or after 2013 can be inferred directly from the conditional probability of having graduated from that HEI between 2000 and 2013. In addition, for obvious reasons, graduates in the LFS data can only have been drawn from the HEMIS graduation distributions for previous years. It is not possible, for example, for a graduate observed in the 2001 March LFS data to only have graduated in 2002. This implies that graduates from the 2000 LFS data could only be probabilistically linked using 2000 HEMIS data, graduates from the 2001 LFS data could only be probabilistically linked using 2000 - 2001 HEMIS data, and so forth.

Under these assumptions, each graduate in the LFS data was probabilistically linked to the HEMIS data. Table

[17] It should be noted that there are reasons to be weary of reporting error on the *"highest level of education completed"* variables in the LFS and QLFS data. This could happen if respondents indicate that they have completed a certain level of education when they have only attended that level without actually completing it. While this is likely to introduce measurement error and may even bias results, it is largely unavoidable given that misreporting errors in the LFS and QLFS data are virtually impossible to detect.

[18] The 2000 to 2007 March and September LFSs use the 12-category South African Qualifications Authority (SAQA) classification of field of study whereas the HEMIS data uses the 22-category Department of Education (DoE) second order classification of educational subject matter (CESM) classification of field of study. In order to use these variables as identifiers in the *p-linking* procedure, it was therefore necessary to convert the 22 CESM fields in the HEMIS data into the 12 SAQA fields as per Mabizela (2005:94).

B.1 shows the number of unique combinations for each of the three linking criteria in the LFS and HEMIS data in relation to the sample sizes for each of the datasets under consideration. Based on this information, Table B.2 in Appendix Appendix B shows the percentages of LFS graduates in each year that could be linked successfully using the available criteria. Once the LFS graduates were linked, the inferred probabilities regarding the specific HEIs from which they are likely to have graduated was used to calculate the respective probabilities that they graduated from a technikon, a comprehensive university, a traditional university, a university of technology, an HDI, an HAI, a Cluster 1 HEI, a Cluster 2 HEI, or a Cluster 3 HEI.

As a further potential diagnostic on the probabilistic linking approach used, Tables B.6, B.7, and B.8 respectively show the actual proportions of HEMIS graduates who graduated from the various types of HEIs listed above, the proportion of graduates in the LFS data sample who, via probabilistic linking, are estimated to have graduated from different types of HEIs, and the proportion of graduates in South Africa's working-age population who are estimated to have graduated from various HEIs.

Lastly, it is important to note that the probabilistic linking approach introduces non-classical measurement error in the estimations that follow. Crucially, the nature of this measurement error differs from that which typically arises in instances where indicator variables are subject to misclassification. Under indicator variable misclassification, measurement error is necessarily correlated with the misclassified indicator variable (Aigner, 1973). That is not the case here. Instead, the measurement error here is akin to measurement error as a result of using group averages to proxy for individual-level variables (Angrist and Krueger, 1999:1342). This is clear when one considers that, as indicated by equation (1), the probabilistic linking approach is effectively tantamount to using group averages (i.e. proportion of graduates who graduated from a specific HEI) from the HEMIS data as proxy variables for missing individual-level HEI indicator variables in the LFS data. By construction, the measurement error will therefore be uncorrelated with the HEI proxy variables. It follows that the parameter estimates on the HEI proxies will be consistent under OLS estimation (Pischke, 2007:9). However, since the HEI proxy variables are imprecisely measured relative to the missing individual-level HEI indicator variables in the LFS data, it is also the case that the standard errors associated with the parameter estimates will be inflated. This is illustrated in greater detail in Appendix B.1.

5.3 The association between HEI type and graduate unemployment/employment

Having assigned to each graduate in the LFS data a set of variables capturing the estimated probability of having graduated from a HEI of specific type, the analysis now proceeds to the estimation of the association between that HEI type and graduate labour market outcomes.

A series of probit regressions were estimated to find the partial association between the probability that a graduate attended a specific type of HEI and the probability that that graduate is (a) narrowly unemployed and (b) employed.[19] Each set of regressions has three permutations. The first uses the same specification in all the regression tables and includes only the main demographic variables that are assumed to have bearing on graduates' probabilities of unemployment/employment in the South Africa.[20] The second permutation includes a specific HEI type probability variable or set of probability variables while the third permutation interacts that HEI type probability variable or set of probability variables with race.

[19] As explained in Section 3, the narrow definition of unemployment is not only the most consistently defined across Stats SA's various labour force surveys, but the difference in broad and narrow unemployment rates for graduates is largely negligible.

[20] All regressions include variables for *age*, *age-squared, race*, *gender*, *level of qualification held*, *province*, *enrolment in education*, and controls for *survey period.*

Each set of results from these estimations is expected to shed light on the following three questions: First, is there a statistically significant association between the probability of having attended a specific type of HEI and the probability of being unemployed or employed? Second, does controlling for the probability of having attended a specific type of HEI change the extent of any unexplained differences in the probability of unemployment or employment between race groups? Finally, does the association between the probability of having attended a specific type of HEI and the probability of unemployment or employment differ across race groups?

5.3.1 HEI type and the expect probability of narrow unemployment for graduates

The results of the various estimations of narrow unemployment probability are presented in Tables C.1 - C.3 in Appendix Appendix C.

Column (1) in Table C.1 confirms most priors regarding the expected relationships between age, race, qualification level and the probability that a graduate will be narrowly unemployed in the South African labour market. It is found that Coloured, Asian, and White graduates are all significantly less likely to be unemployed than their Black counterparts, even once other factors have been taken into account. Similarly, there is a statistically significant negative association between the level of one's graduate qualification and the probability of being unemployed. It is interesting to note, however, that female graduates are statistically no more likely to be unemployed than male graduates.

The estimates in column (2) of Table C.1 show that there is a statistically significant association between the probability of having graduated from a specific type of HEI and the probability of being unemployed. Specifically, graduates who attended traditional universities are found to be statistically significantly less likely to be unemployed than graduates who attended comprehensive universities, but statistically significantly more likely to be unemployed than graduates who attended either technikons or universities of technology. However, the estimates in column (3) show that the extent to which this is true varies by race. For example, Indian graduates from traditional universities are estimated to have lower likelihoods of narrow unemployment than those from technikons or universities of technology. Similarly, the estimated likelihood of narrow unemployment is higher for White graduates from universities of technology that it is for those who attended traditional universities.

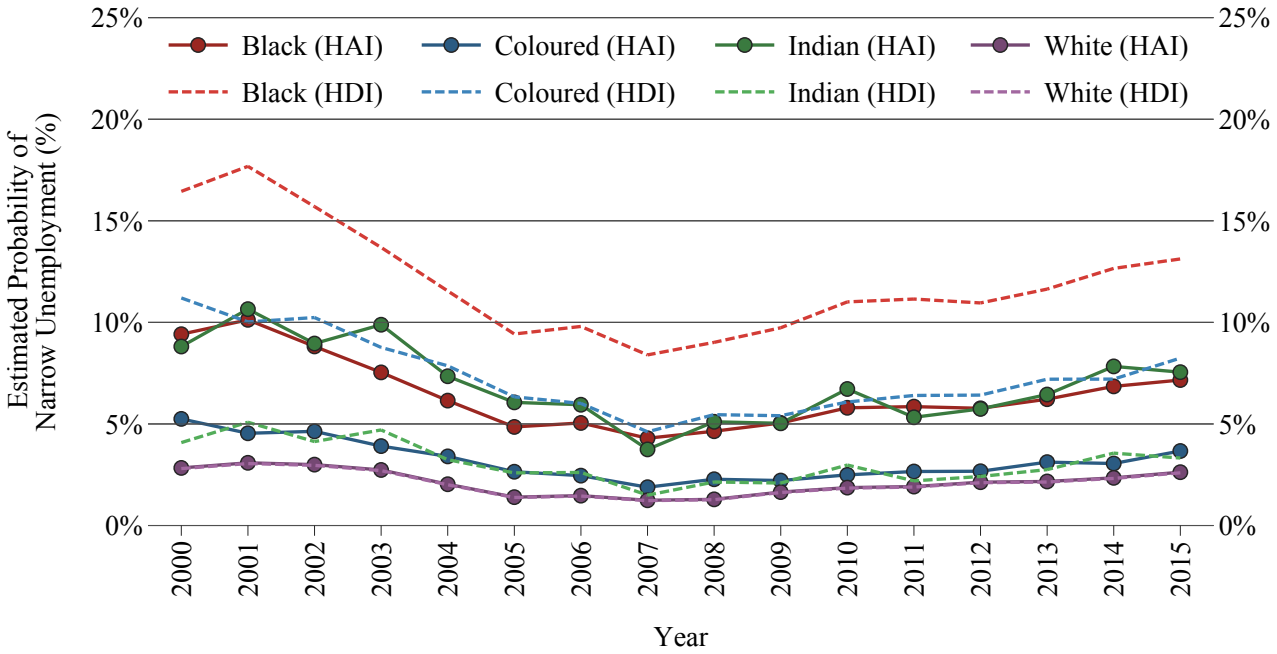
While the fact that the coefficients on the HEI type probability variables in Table C.1 are statistically significant indicates that the type of HEI attended is predictive of the probability of unemployment, the coefficients on the race indicator variables remain statistically significant even after these measures have been taken into account. It follows that the observed racial differentials in graduate unemployment rates cannot be explained away completely by the fact that graduates from different race groups are likely to have graduated from different types of HEI.

Columns (2) and (3) in Table C.2 show not only that graduates who are likely to have graduated from HDIs have statistically significant higher probabilities of being unemployed than their counterparts from HAIs, but that the association between attending an HDI or an HAI and the probability of unemployment also differs between race groups. The coefficients on the interaction terms suggest that the positive association between the likelihood of unemployment and the probability of having graduated from an HDI is effectively negated

for Indian and White graduates.[21] In fact, it would appear as though the probability of unemployment for Indian graduates from HDIs is lower, on average, than it is for those from HAIs. The implication is that the detrimental association between attending an HDI and graduate unemployment appears to apply only to Black and, to a lesser extent, Coloured graduates.

These findings are illustrated in Figure 17 which uses the predictions from regression (3) in Table C.2 to calculate the yearly expected probabilities of narrow unemployment for different race groups, conditional on having graduated either from an HDI or an HAI. Taken in conjunction with the estimates in Table C.2, the graph suggests that part of the unexplained difference in unemployment rates for Black, Coloured, and White graduates can be explained by the fact that Black and Coloured graduates have historically been far more likely to graduate from HDIs than Whites. In fact, the figure shows that, while unexplained differences remain even after controlling for the historical status of the HEI likely to have been attended, the narrow unemployment rates for Black and Coloured graduates from HAIs may be as much as 5 percentage points lower than the narrow unemployment rates for Black and Coloured graduates from HDIs. Nevertheless, it remains clear that the expected level of unemployment among White graduates is still far lower, on average, than it is among Black and Coloured graduates, regardless of the historical status of the HEI attended.

Figure 17: Predicted probability of narrow unemployment for graduates, by HAI/HDI and race (2000 - 2015)



NOTES: Figures reflect the mean predicted graduate narrow unemployment probability for the respective race groups in each year. Predictions based on regression (3) in Table C.2. Estimates of the expected unemployment probability associated with attending a HDI were generated using HDI = 1 and HAI = 0. Estimates of expected unemployment probability associated with attending a HAI were generated using HAI = 1 and HDI = 0. All other variables kept at their observed values in the data when calculating the respective expected graduate unemployment probabilities.

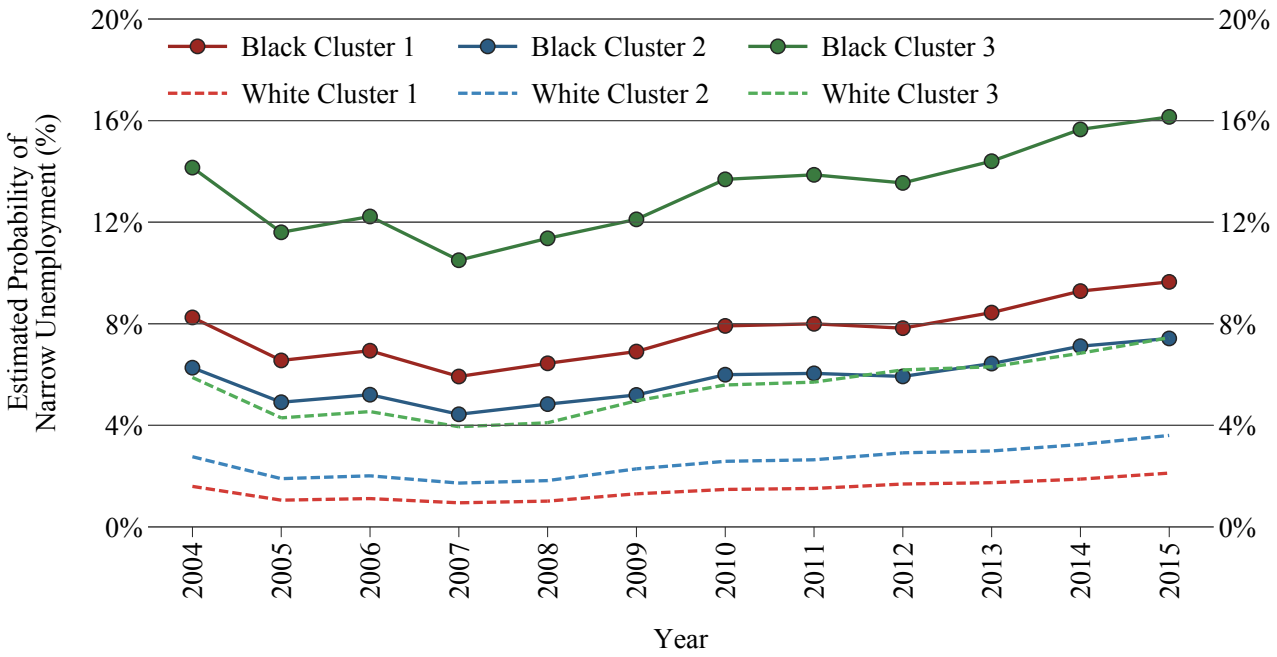
Lastly, the results from regression (2) in Table C.3 suggest that graduates who are likely to have graduated from cluster 2 or cluster 1 HEIs are statistically significantly less likely to be unemployed than graduates from cluster 3 HEIs. As the coefficients on the interactions between the cluster probabilities and the race variables in column (3) of Table C.3 are difficult to interpret, the results for the regression are again graphically illustrated

[21] The statistical insignificance of the interaction term between the *HDI* and *White* variables is likely to be a consequence of the fact that, as discussed above, very few White graduates would have studied at HDIs. It follows that the coefficient in question is imprecisely estimated.

in Figure 18 which plots the expected predicted probabilities of narrow unemployment for Black and White graduates - the two race groups for whom the estimated narrow unemployment rate differentials are largest - conditional on the cluster of the HEI attended.

It is clear that the predicted probability of unemployment within each race group is far higher for graduates from cluster 3 institutions than those from cluster 1 and cluster 2 HEIs. However, it is telling that the expected narrow unemployment rate for Black graduates from cluster 1 and 2 HEIs is at least as high as the expected narrow unemployment rate for White graduates from cluster 3 HEIs. When viewed purely in terms of the probability of unemployment, these results thus suggest that the best-performing group of Black graduates only performs the same as, if not worse than, the worst-performing group of White graduates.

Figure 18: Predicted probability of narrow unemployment for Black and White graduates, by HEI Cluster (2004 - 2015)



NOTES: Figures reflect the mean predicted graduate narrow unemployment probability for the respective race groups in each year. Predictions based on regression (3) in Table C.3. Estimates correspond to the mean predicted narrow unemployment probability for the respective race groups in each year. Estimates of expected graduate unemployment probability associated with attending a cluster 1 HEI generated using cluster1 = 1, cluster2 = 0, cluster3 = 0; estimates associated with attending a cluster 2 HEI generated using cluster1 = 0, cluster2 = 1, cluster3 = 0; estimates associated with attending a cluster 0 HEI generated using cluster1 = 0, cluster2 = 0, cluster3 = 1. All other variables kept at their observed values in the data when calculating the respective expected graduate unemployment probabilities.

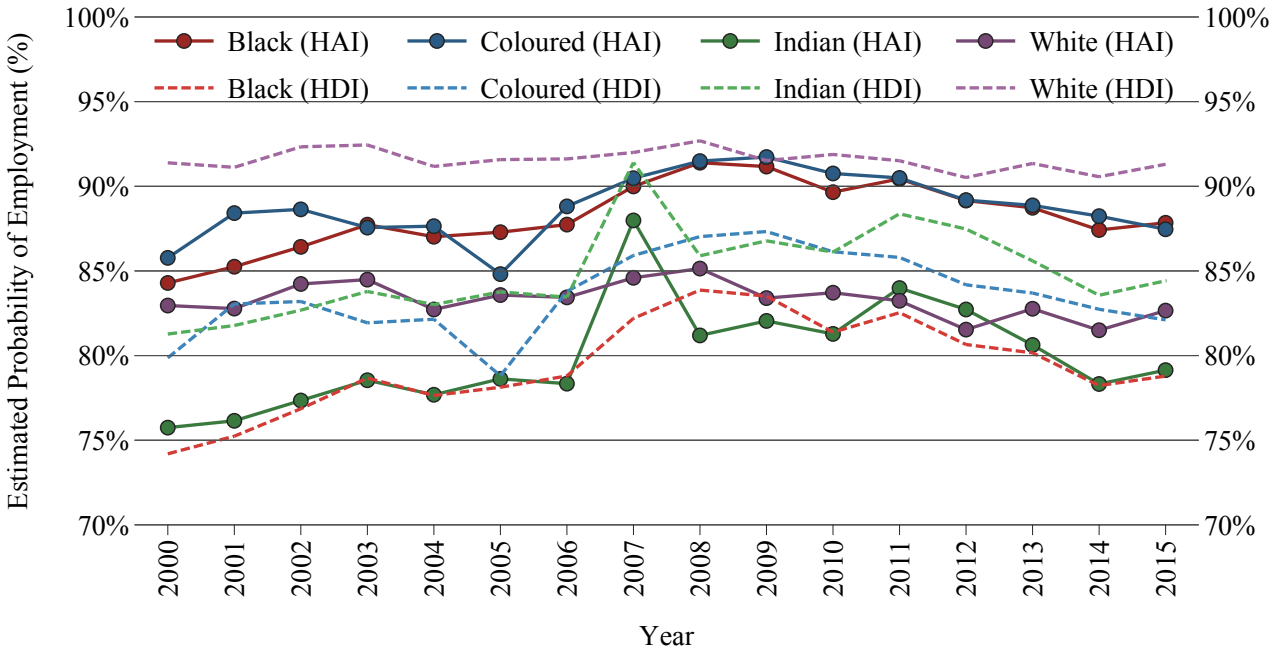
5.3.2 HEI type and the expect probability of employment for graduates

Tables C.4 - C.6 contain the results from probit regressions estimating the probability of graduate employment. The specifications used are the same as in Tables C.1 - C.3 with the exception that, in this instance, the dependent variable is employment status rather than narrow unemployment status.

The results in Table C.4 are broadly consistent with those presented in Tables C.1 - C.3 in terms of the nature of the conditional associations that are found to exist between specific demographic variables and graduate labour market status. However, there are some noteworthy differences. Chief among these is the fact that female graduates are found to have statistically significant lower probabilities of employment than male

graduates. Given that the narrow unemployment rate is simply the difference between the narrow LFP rate and the employment rate, this finding implies that the fact the the narrow unemployment rates for female graduates are not statistically significantly different from those for male graduates can largely be attributed to the fact that female graduates have lower narrow LFP rates, on average, than male graduates do. A similar argument is likely to explain why Indian graduates are found to be statistically significantly less likely to be employed than Black or Coloured graduates, while simultaneously having been found to be statistically significantly less likely to be unemployed than these two groups.[22]

Figure 19: Predicted probability of employment for graduates, by HAI/HDI and race (2000 - 2015)



NOTES: Figures reflect the mean predicted graduate employment probability for the respective race groups in each year. Predictions based on regression (3) in Table C.5. Estimates of the expected employment probability associated with attending a HDI were generated using HDI = 1 and HAI = 0. Estimates of expected employment probability associated with attending a HAI were generated using HAI = 1 and HDI = 0. All other variables kept at their observed values in the data when calculating the respective expected graduate employment probabilities.

Unsurprisingly, the estimates in column (2) and (3) of Table C.4 suggest that the expected employment rates are highest for graduates from technikons and universities of technology, and lowest for graduates from comprehensive universities, on average and with all else being constant. However, none of the interaction effects between race and the HEI type probability variables in column (3) are statistically significant. Thus, there do not appear to be any significant differences in the associations between HEI type and employment probabilities for graduates from different race groups. Moreover, much like the case for the narrow unemployment estimations, controlling for the likelihood of having graduated from a specific type of HEI does not greatly alter the statistical significance or apparent magnitude of the unexplained racial employment rate differentials for graduates.

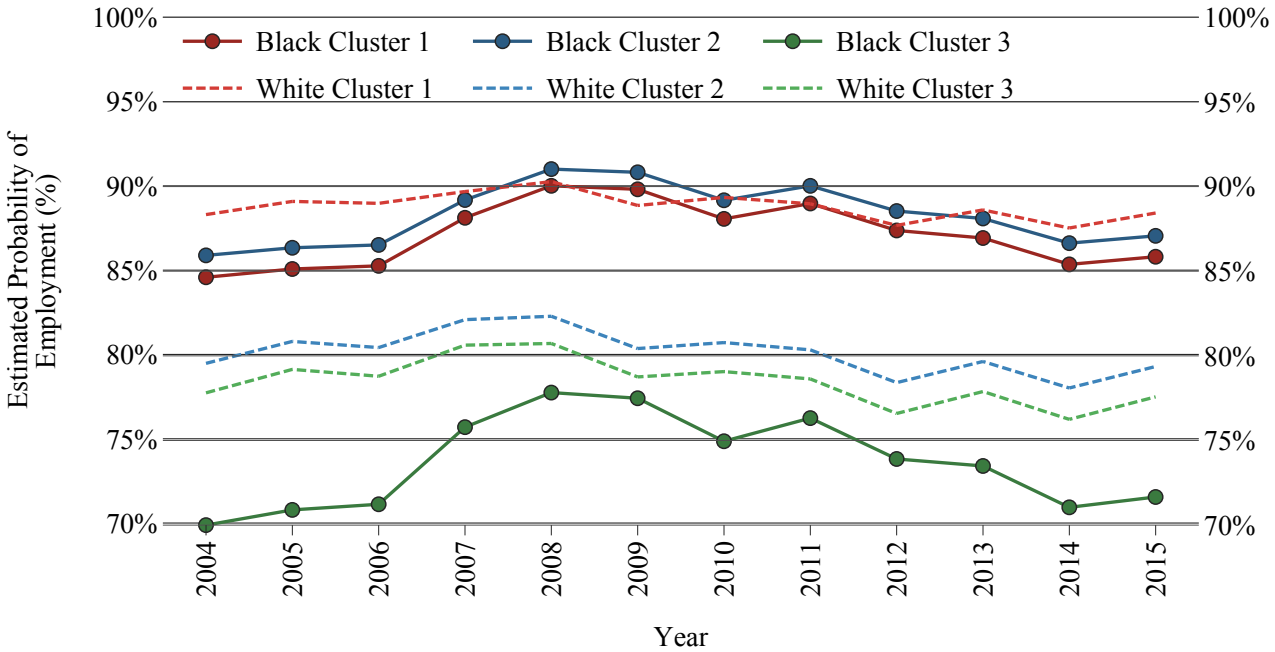
The results from column (2) in Table C.5 show that the coefficient on the *White* race indicator variable becomes statistically insignificant once a measure of the historical status of the HEI from which graduates are likely to have graduated is included in the estimation. However, the HDI probability variable is not statist-

[22] The LFS and QLFS data support this notion. Over the period 2000 - 2015, the average LFP rate for male graduates was 92.8% whereas, for females, it was only 85.1%. Similarly, the the average LFP rate for Indian graduates (86.4%) was lower than the average LFP rates for Black (93.1%) and Coloured (91.1%) graduates.

ically significant in column (2). Though this may appear perplexing given the findings discussed in Section 5.3.1 above, the reason or the statistical insignificance of the *HDI* term in column (2) becomes clear when it is interacted with the respective race variables in column (3) of Table C.5. These estimates show that the nature and extent of the association between the probability of having graduated from an HDI and the likelihood of employment differs substantially across race groups. Specifically, the results suggest that there is a large and statically significant negative association between graduating from an HDI and the probability of employment for Black and Coloured graduates. By contrast, once other factors have been taken into account, the association between graduating from an HDI and the probability of employment for White and Indian graduates might actually be positive.

As before, these findings are illustrated graphically in Figure 19. The graph shows that, while the expected probability of employment for Black graduates from HDIs are the lowest for all groups, the expected probability of employment for Black or Coloured graduates from HAIs is higher than that for White graduates from HAIs. However, it is important to note that the difference between the expected employment probabilities for Black and White graduates from HAIs is smaller than the difference between the expected employment probabilities for Black and White graduates from HDIs. This finding resonates with those found by others on the primary and secondary schooling system in South Africa which suggest that, while Black individuals may benefit from attending HAIs, this benefit is generally overshadowed by the significant disadvantages associated with attending HDIs (van der Berg, 2007; Van der Berg, 2008).

Figure 20: Predicted probability of employment for Black and White graduates, by HEI Cluster (2004 - 2015)



NOTES: Figures reflect the mean predicted graduate narrow employment probability for the respective race groups in each year. Predictions based on regression (3) in Table C.3. Estimates correspond to the mean predicted employment probability for the respective race groups in each year. Estimates of expected graduate employment probability associated with attending a cluster 1 HEI generated using cluster1 = 1, cluster2 = 0, cluster3 = 0; estimates associated with attending a cluster 2 HEI generated using cluster1 = 0, cluster2 = 1, cluster3 = 0; estimates associated with attending a cluster 0 HEI generated using cluster1 = 0, cluster2 = 0, cluster3 = 1. All other variables kept at their observed values in the data when calculating the respective expected graduate employment probabilities.

The results from Table C.6 suggest that the cluster of the HEI likely to have been attended has important bearing on the probability of employment. As before, the predictions from regression (3) are used to calculate the expected employment probabilities from Black and White graduates from different cluster HEIs in Figure

20. The graph shows that, while Black graduates from cluster 3 HEIs are predicted to have the lowest expected probabilities of employment, the probability of employment for Black graduates from cluster 2 or cluster 1 HEIs has been more-or-less the same as the probability of employment for White graduates from cluster 1 HEIs since 2007. Similar to what was the case when looking at narrow unemployment probabilities in Table C.3, White graduates from cluster 2 or cluster 3 institutions have significantly lower employment probabilities than White graduates from cluster 1 HEIs.

Table 1: Predicted average narrow unemployment and employment rates (%) for graduates by race and HEI type (2000 - 2015)

	Narrow Unemployment Rate (%)[a]				*Employment Rate (%)*[b]			
HEI Type	**Black**	**Coloured**	**Indian**	**White**	**Black**	**Coloured**	**Indian**	**White**
Traditional	8.1***	3.1***	2.6***	1.7***	85.0***	88.9***	84.4***	86.8***
Technikon	5.7***	1.0*	4.4**	1.3***	89.9***	91.4***	87.4***	88.4***
Technology	2.4***	3.1	7.4*	2.6	91.6***	93.9***	87.4***	93.1***
Comprehensive	16.2***	10.3*	22.7***	5.7***	68.6***	78.0***	61.5***	72.3***
HDI	11.4***	5.4***	2.6***	2.0**	78.7***	85.9***	84.9***	92.3***
HAI	6.0***	2.2***	6.0***	2.0***	87.6***	90.5***	79.9***	84.4***
Cluster 1	7.7***	3.5***	2.5***	1.5***	86.0***	89.9***	84.5***	89.8***
Cluster 2	5.9***	3.3***	8.6***	2.6***	87.2***	88.2***	76.8***	81.5***
Cluster 3	13.4***	0.0	0.0	5.5	72.1***	72.5	100.0***	79.9***

NOTES: [a]Figures reflect the average predicted graduate narrow unemployment rates for the respective race groups and HEI types over the period 2000 - 2015 and are based on the marginal predictions from the regressions in columns (3) of Tables C.1 - C.3. [b]Figures reflect the average predicted graduate employment rates for the respective race groups and HEI types over the period 2000 - 2015 and are based on the marginal predictions from the regressions in columns (3) of Tables C.4 - C.6. Predictions generated by setting the relevant HEI type proxy variables equal to 1 or 0. E.g. the predicted rates for graduates from traditional HEIs was generated using Technikon = 0, Technology = 0, and Comprehensive = 0, whereas the predicted rates for graduates from Cluster 2 HEIs was generated using Cluster 1 = 0 and Cluster 2 = 1.All other variables were kept at their observed values in the data when calculating the respective expected graduate unemployment/employment rates. *Significant at the 10% level **Significant at the 5% level *** Significant at the 1% level. Significance levels are based on linearised robust standard errors which have been adjusted for complex survey design. Estimates are weighted.

Lastly, Table 1 summarizes the predicted average narrow unemployment rates and employment rates for graduates by race and HEI type over the period 2000 - 2015, using the estimation results from regression (3) in Tables C.1 - C.6. The significant variation in the estimates across race groups and HEI types in the Table provide a clear illustration of the degree to which graduate unemployment and employment rates differ between different groups of graduates in South Africa.

6 Conclusion

The apparent paradox of high levels of graduate unemployment combined with persistent skills shortages in the South African labour market has often been attributed to structural changes which are held to have resulted in a misalignment between the skills that graduates traditionally have to offer and the skills that employers demand. It is claimed that the effects of this supposed skills-mismatch are further exacerbated by the severe heterogeneity in the quality of education received, even at the tertiary level, by different groups and cohorts in South Africa. When coupled with the signal eroding effect of substantial qualification inflation in the labour force over time and the significant changes in the demographic composition of South Africa's stock of graduates, it seems likely that this heterogeneity will have served to undermine the fidelity of graduate

education credentials as signals of potential labour market productivity and, in general, reduced graduate employability.

The results from this paper suggest that graduate unemployment in South Africa is not nearly as problematic as is often asserted. In part, this is simply because individuals with degrees or higher qualifications are often misguidedly lumped together with individuals with post-secondary diplomas and certificates under the collective "graduates". Yet, the descriptive analysis in Section 2 shows precisely why such practice is dubious and leads to an inflated perception of graduate unemployment in South Africa.

Despite significant changes in the demographic composition of South Africa's stock of graduates and policy changes which have altered South Africa's HE landscape, graduates remain the group with the best labour market prospects relative to other education cohorts. This is true for all race groups, even though there remain differences in the employment and unemployment probabilities for Black, Coloured, Indian and White graduates. However, as the multivariate analysis shows, part of the racial differentials in graduate unemployment and employment outcomes in the country can potentially be attributed to heterogeneity in the types of HEIs commonly attended by individuals from different race groups. For example, it is clear that having attended an HDI rather than an HAI is negatively associated with employment prospects and positively linked to the probability of unemployment. Similarly, graduates from Cluster 1 HEIs appear to have higher employment rates and lower unemployment rates than graduates from Cluster 2 or Cluster 3 HEIs.

It is important to note that these findings cannot make any causal claims regarding the relationships between HEI-type and graduate labour market outcomes. It is not, for example, argued that the fact that graduates from HDIs appear to have higher unemployment rates than graduates from HAIs is a consequence of the fact that the quality of education at HDIs is lower than the quality of education at HAIs. While such an argument may be plausible, it is only one of many plausible reasons that may explain the observed associations between HEI type and graduate unemployment/employment rates. As is discussed in the next paper, selection into HE and HEIs is an endogenous process and individuals who graduate from HDIs may be fundamentally different from those who graduate from HAIs in ways that are not accounted for in the estimations presented here.

In addition to the fact that the estimates presented in this paper cannot be interpreted causally, it should also be remembered that the probabilistic linking approach underlying those estimates is based on a number of potentially contestable assumptions. The consistency of the parameter estimates discussed above is ultimately premised on the validity of these assumptions. Thus, while the probabilistic linking methodology offers a novel way of linking HEI aspects to graduate labour market outcomes, it is not without potential flaws.

Notwithstanding the aforementioned caveats, the findings from the analysis suggest that understanding the heterogeneity between HEIs may be crucial for understanding the observed variation in graduate labour market outcomes as well as the racial differentials in graduate unemployment rates. Consequently, more should be done to ensure that HEI-related factors are incorporated when analysing graduate labour market prospects and it is essential for researchers to have access to the type of data that would enable them to do so. Ultimately, policy interventions aimed at improving graduate labour market outcomes can only be effective if the nature of the racial and institutional dimensions underlying those outcomes are understood.

References

Aigner, D. J. (1973). Regression with a binary independent variable subject to errors of observation. *Journal of Econometrics* **1**(1), pp. 49 – 60.

Altman, M. (2007). *Youth Labour Market Challenges in South Africa*. Pretoria: Human Sciences Research Council.

Angrist, J. D. and Krueger, A. B. (1999). Empirical Strategies in Labor Economics. In O. Ashenfelter and D. E. Card (Eds.) *Handbook of Labor Economics*. vol. 3. chap. 23, pp. 1277 – 1366. Elsevier BV. 1st edn.

Baldry, K. (2015). Graduate unemployment in South Africa: social inequality reproduced. *Journal of Education and Work* **2015**, pp. 1 – 25.

Bhorat, H. (2004). Labour market challenges in post-Apartheid South Africa. *South African Journal of Economics* **72**(5), pp. 940 – 977.

Bhorat, H. and Mayet, N. (2012). Labour Demand Trends and The Determinants of Unemployment in South Africa. Closing the Skills and Technology Gap in South Africa: Background Paper 4. World Bank Human Development Group, Africa Region.

Bhorat, H., Mayet, N. and Visser, M. (2010). Student graduation, labour market destinations and employment earnings. In M. Letseka, M. Cosser, M. Breier and M. Visser (Eds.) *Student Retention & Graduate Destination: Higher Education & Labour Market Access & Success*. chap. 7, pp. 97 – 124. Pretoria: HSRC Press.

Blom, R. (2011). The size and shape of private, post-school education and training provision in South Africa: A study for the Department of Higher Education and Training. Johannesburg: Centre for Education Policy Development (CEPD).

Branson, N., Leibbrandt, M. and Zuze, T. L. (2009*a*). The Demand for Tertiary Education in South Africa. Final report to the centre for higher education transformation (chet). Southern Africa Labour and Development Research Unit (SALDRU), University of Cape Town (UCT).

——— (2009*b*). What are the returns of tertiary education and who benefits? In N. Cloete (Ed.) *Responding to the Educational Needs of Post-School Youth: Determining the Scope of the Problem and Developing a Capacity-Building Model*. chap. 3, pp. 45 – 60. Centre for Higher Education Transformation (CHET) / African Minds. 1st edn.

Bunting, I. (2002). The higher education landscape under apartheid. In N. Cloete, R. fehnel, P. Maasen, T. Moja, H. Perold and T. Gibbon (Eds.) *Transformation in Higher Education: Global Pressures and Local Realities in South Africa*. pp. 35 – 52. Netherlands: Kluwer Academic Publishers. 2 edn.

CHE (2004). *South African Higher Education in the first decade after democracy*. Pretoria: Council on Higher Education.

——— (2010). Access and throughput in South African Higher Education: Three case studies. HE Monitor No. 9. Pretoria: Council on Higher Education.

CHEC (2013). *Pathways from University to Work: A Graduate Destination Survey of the 2010 Cohort of Graduates from the Western Cape Universities*. Wynberg: Cape Higher Education Consortium (CHEC).

CHET (2010). Clustering of South African universities into distinct groups based on input and output indicators. Unpublished report. Wynberg: Centre for Higher Education Transformation.

DHET (2014). Graduates Table 2.13 for all institutions. Excel Tables. [online] available from http://www.dhet.gov.za/HEMIS/Graduates/ [Accessed 24 July 2014].

——— (2015). Statistics on Post-School Education and Training in Sounth Africa: 2013. Report. Department of Higher Education and Training.

DRPU (2006). Graduate Unemployment in Post-Apartheid South Africa: Nature and Possible Policy Responses. Research Report Compiled for Business Leadership South Africa. Cape Town: University of Cape Town, Development Policy Reseach Unit.

Financial and Fiscal Commission (2012). Technical Report: Submission for the 2013/14 Division of Revenue. Technical report. Financial and Fiscal Commission.

Fisher, G. and Scott, I. (2011). The Role of Higher Education in Closing the Skills Gap in South Africa. Closing the Skills and Technology Gap in South Africa: Background Paper 3. World Bank Human Development Group, Africa Region.

Goldstein, H., Harron, K. and Wade, A. (2012). The analysis of record-linked data using multiple imputation with data value priors. *Statististics in Medicine* **31**(28), pp. 3481 – 3493.

HEDA (2015). Higher Education Data Analyzer. [online] available from http://www.heda.co.za/pds/background.aspx [Accessed September 11, 2015].

Hof, M. H. P. and Zwinderman, A. H. (2012). Methods for analyzing data from probabilistic linkage strategies based on partially identifying variables. *Statististics in Medicine* **31**(30), pp. 4231 – 4242.

Humburg, M., De Grip, A. and Van der Velden, R. (2012). Which skills protect graduates against a slack labour market? Working Paper RM/12/002. Maastricht University School of Business and economics.

Kim, G. and Chambers, R. (2012). Regression analysis under incomplete linkage. *Computational Statistics & Data Analysis* **56**(9), pp. 2756 – 2770.

Kingdon, G. and Knight, J. (2006). The measurement of unemployment when unemployment is high. *Labour Economics* **13**(3), pp. 291 – 315.

Koen, C. (2006). *Higher education and work: Setting a new research agenda.* Cape Town: HSRC Press.

Kraak, A. (2010). The collapse of the graduate labour market in South Africa: evidence from recent studies. *Research in Post-Compulsory Education* **15**(1), pp. 81 – 102.

——— (2015). The value of graduate destination survey data in understanding graduate unemployment: A focus on the universities of technology. *South African Journal of Labour Relations* **39**(1), pp. 93 – 113.

Kruss, G. (2007). Employment and employability: expectations of higher education responsiveness in South Africa. *Journal of Education Policy* **19**(6), pp. 673 – 689.

Mabizela, M. (2005). *The Business of Higher Education - A study of public-private partnerships in the provision of Higher Education in South Africa.* Pretoria: HSRC Press.

Maharasoa, M. and Hay, D. (2010). Higher Education and Graduate Employment in South Africa. *Quality in Higher Education* **7**(2), pp. 139 – 147.

Mlatsheni, C. and Rospabe, S. (2002). Why is Youth Unemployment so High and Unequally spread in South Africa? DPRU Working Paper 02/65. Development Policy Research Unit, University of Cape Town.

Moleke, P. (2010). The Gradaute Labour Market. In M. Letseka, M. Cosser, M. Breier and M. Visser (Eds.) *Student Retention & Graduate Destination: Higher Education & Labour Market Access & Success*. chap. 6, pp. 87 – 95. Pretoria: HSRC Press.

Naong, M. N. (2011). Promotion of Entrepreneurship Education - A Remedy to Graduates and Youth Unemployment - A Theoretical Perspective. *Journal of Social Science* **28**(3), pp. 181 – 189.

NPC (2011). National Development Plan: Vision for 2030. Pretoria: National Planning Commission.

Nunez, I. and Livanos, I. (2010). Higher education and unemployment in Europe: an analysis of the academic subject and national effects. *Higher Education* **59**(4), pp. 475 – 487.

Pauw, K., Oosthuizen, M. and van der Westhuizen, C. (2008). Graduate Unemployment in the Face of Skills Shortages: A Labour Market Paradox. *South African Journal of Economics* **76**(1), pp. 45–57.

Pischke, S. (2007). Lecture Notes on Measurement Error. [online] available from http://econ.lse.ac.uk/staff/spischke/ec524/Merr_new.pdf [Accessed July 4, 2015].

Ridder, G. and Moffitt, R. (2007). The Econometrics of Data Combination. In J. J. Heckman and E. E. Leamer (Eds.) *Handbook of Econometrics*. chap. 75, pp. 5469 – 5547. Elsevier BV.

Scott, I., Yeld, N. and Hendry, J. (2007). Higher Education Monitor: A Case for Improving Teaching and Learning in South African Higher Education. Cape Town: Council on Higher Education (CHE).

Teichler, U. (2007). Does Higher Education Matter? Lessons from a Comparative Graduate Survey. *European Journal of Education* **42**(1), pp. 11 – 34.

van der Berg, S. (2007). Apartheid's Enduring Legacy: Inequalities in Education. *Journal of African Economies* **16**(5), pp. 849 – 880.

Van der Berg, S. (2008). How effective are poor schools? Poverty and educational outcomes in South Africa. *Studies in Educational Evaluation* **34**(3), pp. 145 – 154.

Van der Merwe, A. (2010). Does Human Capital Theory Explain the Value of Higher Education? A South African Case Study. *American Journal of Business Education* **3**(1), pp. 107 – 118.

Wu, C.-C. (2011). High graduate unemployment rate and Taiwanese undergraduate education. *International Journal of Educational Development* **31**(3), pp. 303 – 310.

Yu, D. (2008). The South African labour market: 1995 - 2006. Stellenbosch Economic Working Paper 05/08. Department of Economics and the Bereau for Economic Research at the University of Stellenbosch.

——— (2010). Defining and measuring informal employment in South Africa: A review of recent approaches. Stellenbosch Economic Working Paper 09/10. Department of Economics and the Bereau for Economic Research at the University of Stellenbosch.

Appendix A HEI and HE qualification classifications

Table A.1: NQF exit levels of HEMIS qualification types

	Qualification type	Code	NQF Exit Level
Technikons (1986 - 2003)	National Certificate	21	5
	National Higher Certificate	22	6
	National Diploma	23	6
	National Higher Diploma	25	7
	Baccalaureus Technologiae Degree	26	7
	Post-diploma Diploma	24	8
	Master's Diploma in Technology	27	9
	Magister Technologiae Degree	28	9
	Laureatus in Technology	29	10
	Doctor Technologiae Degree	30	10
Universities (1986 - 2003)	Undergraduate Diploma or Certificate (1 or 2 years)	11	5
	Undergraduate Diploma or Certificate (3 yrs)	1	6
	General Academic Bachelor's Degree	2	7
	Professional First Bachelor's Degree (3 years)	33	7
	Professional First Bachelor's Degree (4 years)	3	8
	Post-graduate Diploma or Certificate	4	8
	Post-graduate Bachelor's Degree	5	8
	Honours Degree	6	8
	Masters Degree	7	9
	Doctoral Degree	8	10
All HEIs (2004 -)	Higher Certificate	41	5
	Advanced Certificate	42	6
	Diploma	43	6
	Advanced Diploma	44	7
	Bachelor's Degree (360 credits)	45	7
	Bachelor's Degree (480 credits)	46	8
	Postgraduate Diploma	47	8
	Bachelor Honours Degree	48	8
	Master's Degree	49	9
	Doctoral Degree	50	10

NOTES: National Qualifications Framework (NQF) exit level classifications of Higher Education Management Information System (HEMIS) qualifications based on the Higher Education Qualifications Framework (HEQF) Implementation Template and the South African Qualifications Authority's (SAQA) suggested NQF exit level classifications.

Table A.2: HEI classifications, before and after amalgamation in 2004

	Pre-Amalgamation			Post-Amalgamation		
	Instcode	**Institution**	**Type**	**Instcode**	**Institution**	**Type**
Former traditional universities (before 2004)	101	University of Cape Town	HAI	H02	UCT	HAI
	102	University of Durban-Westville	HDI	H08	UKZN	HDI
	103	University of Fort Hare	HDI	H05	UFH	HDI
	104	Medical Uni. of South Africa	HDI	H09	UL	HDI
	105	University of Natal	HAI	H08	UKZN	HDI
	106	University of the North	HDI	H09	UL	HDI
	107	University of the Free State	HAI	H06	UFS	HAI
	108	University of Port Elizabeth	HAI	H10	NMMU	HAI
	109	Potchefstroom University	HAI	H11	NWU	HDI
	110	University of Pretoria	HAI	H12	UP	HAI
	111	Rand Afrikaans University	HAI	H07	UJ	HAI
	112	Rhodes University	HAI	H13	RU	HAI
	113	University of South Africa	HAI	H14	UNISA	HAI
	114	University of Stellenbosch	HAI	H15	US	HAI
	115	University of Western Cape	HDI	H20	UWC	HDI
	116	University of Witwatersrand	HAI	H21	WITS	HAI
	117	University of Zululand	HDI	H22	UZ	HDI
	118	Vista University	HDI	—	—	—
	119	University of Transkei	HDI	H19	WSU	HDI
	120	University of North West	HDI	H11	NWU	HDI
	121	University of Venda	HDI	H17	UNIVEN	HDI
Former technikons (before 2004)	301	Cape Technikon	HAI	H01	CPUT	HAI
	302	Northern Gauteng Technikon	HDI	H16	TUT	HDI
	303	Mangosuthu Technikon	HDI	H25	MUT	HDI
	304	M.L. Sultan Technikon	HDI	H04	DUT	HDI
	305	Natal Technikon	HAI	H04	DUT	HDI
	306	Technikon Free State	HAI	H03	CUT	HAI
	307	Peninsula Technikon	HDI	H01	CPUT	HDI
	308	Port Elizabeth Technikon	HAI	H10	NMMU	HAI
	309	Pretoria Technikon	HAI	H16	TUT	HDI
	310	Technikon SA	HAI	H14	UNISA	HAI
	311	Vaal Triangle Technikon	HAI	H18	VUT	HAI
	312	Witwatersrand Technikon	HAI	H07	UJ	HAI
	313	Border Technikon	HDI	H19	WSU	HDI
	314	Technikon North West	HDI	H16	TUT	HDI
	315	Eastern Cape Technikon	HDI	H19	WSU	HDI

NOTES: Former and current HE institution classifications based on Financial and Fiscal Commission (2012:55), Bunting (2002:49 - 51). * Vista University's satellite campuses were merged in to various universities including Nelson Mandela Metropolitan University (NMMU), University of the Free State (UFS), University of Johannesburg (UJ), University of Pretoria (UP), University of South Africa (UNISA), and the Vaal University of Technology (VUT).

Table A.3: Current HEI classifications

Instcode	Abbreviation	Institution	Type	HAI/HDI	CHET Cluster
H01	CPUT	Cape Peninsula University of Technology	Technology	HDI	2
H02	UCT	University of Cape Town	Traditional	HAI	1
H03	CUT	Central University of Technology, Free State	Technology	HAI	2
H04	DUT	Durban Institute of Technology	Technology	HDI	2
H05	UFH	University of Fort Hare	Traditional	HDI	2
H06	UFS	University of the Free State	Traditional	HAI	2
H07	UJ	University of Johannesburg	Comprehensive	HAI	2
H08	UKZN	University of KwaZulu-Natal	Traditional	HDI	1
H09	UL	University of Limpopo	Traditional	HDI	3
H10	NMMU	Nelson Mandela Metropolitan University	Comprehensive	HAI	2
H11	NWU	North West University	Traditional	HDI	1
H12	UP	University of Pretoria	Traditional	HAI	1
H13	RU	Rhodes University	Traditional	HAI	1
H14	UNISA	University of South Africa	Comprehensive	HAI	2
H15	US	University of Stellenbosch	Traditional	HAI	1
H16	TUT	Tshwane University of Technology	Technology	HDI	2
H17	UV	University of Venda	Comprehensive	HDI	3
H18	VUT	Vaal University of Technology	Technology	HAI	3
H19	WSU	Walter Sisulu University	Comprehensive	HDI	3
H20	UWC	University of Western Cape	Traditional	HDI	2
H21	WITS	University of Witwatersrand	Traditional	HAI	1
H22	UZ	University of Zululand	Comprehensive	HDI	3
H25	MUT	Mangosuthu University of Technology	Technology	HDI	3

NOTES: Current HE institution classifications based on Financial and Fiscal Commission (2012:55), Bunting (2002:49 - 51). * Vista University's various satellite campuses were merged in to various universities including Nelson Mandela Metrolpolitan University (NMMU), University of the Free State (UFS), University of Johannesburg (UJ), University of Pretoria (UP), University of South Africa (UNISA), and the Vaal University of Technology (VUT). Traditional: Traditional Unvirsity; Comprehensive: Comprehensive University; Technology: University of Technology; HDI: Historically Disadvantaged Institution; HAI: Historically Advantaged Institution; Cluster1: CHET Cluster 1

Appendix B HEMIS - LFS/QLFS probabilistic linking diagnostics

Table B.1: Unique combinations and sample sizes across probabilistic linking criteria for the LFS and HEMIS samples

LFS / QLFS Years	HEMIS Data Years	Unique Combinations of Criteria Variables						Sample Size	
		Criteria 1		Criteria 2		Criteria 3			
		LFS	*HEMIS*	*LFS*	*HEMIS*	*LFS*	*HEMIS*	*LFS*	*HEMIS*
2000	2000	1 078	6 236	311	1 291	318	422	2 172	62 252
2001	2000 - 2001	1 373	8 044	332	1 426	348	449	3 458	123 232
2002	2000 - 2002	1 355	9 170	320	1 514	327	474	3 557	186 674
2003	2000 - 2003	1 332	10 024	321	1 597	326	492	3 478	254 398
2004	2000 - 2004	2 005	10 791	869	1 670	328	514	3 120	327 076
2005	2000 - 2005	1 738	11 457	808	1 734	332	524	2 722	399 979
2006	2000 - 2006	1 677	12 026	779	1 788	311	537	2 620	474 238
2007	2000 - 2007	1 648	12 590	764	1 850	317	552	2 598	549 793
2008	2000 - 2008	—	13 134	1 006	1 912	344	563	6 690	628 913
2009	2000 - 2009	—	13 644	998	1 958	343	574	6 412	713 602
2010	2000 - 2010	—	14 416	961	2 022	340	588	6 240	805 398
2011	2000 - 2011	—	15 085	1 037	2 073	347	598	6 693	902 738
2012	2000 - 2012	1 749	15 730	1 042	2 121	344	605	6 532	1 007 976
2013	2000 - 2013	2 990	16 277	1 055	2 165	351	617	7 089	1 122 981
2014	2000 - 2013	2 957	16 277	1 063	2 165	351	617	7 230	1 122 981
2015	2000 - 2013	1 882	16 277	857	2 165	327	617	3 292	1 122 981

NOTES: Figures represent (a) the number of unique combinations of variables for each of the match criteria used and (b) the sample sizes of the respective LFS and corresponding HEMIS data samples against which they were probabilistically matched. Samples included only graduates with NQF exit level 7 or higher qualifications. Criterion 1: Unique combination of year of birth, race, gender, level of degree awarded (bachelor, postgraduate, etc), and SAQA field of study. Criterion 2: Unique combination of year of birth, race, gender, and level of degree awarded (bachelor, postgraduate, etc). Criterion 3: Unique combination of year of birth, race, and gender. Given that the "field of study" variable was not asked in the 2008 - 2011 QLFS questionnaires, it was not possible to use criterion 1 to probabilistically matched 2008 - 2011 LFS graduates to HEMIS data. Figures correspond to sample estimates and are unweighted.

Table B.2: Percentage of LFS/QLFS sample graduates probabilistically linked to HEMIS data, by criterion

LFS / QLFS Years	HEMIS Years to Match on	Percentage successfully 'linked' (%)			
		Criteria 1	Criteria 2	Criteria 3	Unmatched
2000	2000	80.0	16.9	1.8	1.2
2001	2000 - 2001	88.4	10.2	0.4	1.0
2002	2000 - 2002	90.2	9.0	0.4	0.3
2003	2000 - 2003	91.5	7.9	0.3	0.3
2004	2000 - 2004	91.3	8.0	0.2	0.4
2005	2000 - 2005	92.0	7.3	0.1	0.6
2006	2000 - 2006	92.1	7.6	0.1	0.1
2007	2000 - 2007	91.7	7.5	0.2	0.7
2008	2000 - 2008	0.0	100.0	0.0	0.0
2009	2000 - 2009	0.0	100.0	0.0	0.0
2010	2000 - 2010	0.0	100.0	0.0	0.0
2011	2000 - 2011	0.0	100.0	0.0	0.0
2012	2000 - 2012	44.0	55.9	0.0	0.0
2013	2000 - 2013	94.3	5.6	0.0	0.1
2014	2000 - 2013	93.8	6.1	0.1	0.1
2015	2000 - 2013	93.4	6.4	0.1	0.0

NOTES: Figures represent the percentages of graduates for each year of the pooled LFS/QLFS sample that were probabilistically matched using a specific criterion. Linking criteria were used sequentially: An a attempt was made to probabilistically match on criterion 1 first, then on criterion 2 and, finally, on criterion 3. The LFS/QLFS sample included only graduates with NQF exit level 7 or higher qualifications. Criterion 1: Unique combination of year of birth, race, gender, level of degree awarded (bachelor, postgraduate, etc), and SAQA field of study. Criterion 2: Unique combination of year of birth, race, gender, and level of degree awarded (bachelor, postgrad, etc). Criterion 3: Unique combination of year of birth, race, and gender. Given that the "field of study" variable was not asked in the 2008Q1 - 2012Q2 QLFS questionnaires, it was not possible to use criterion 1 to probabilistically match 2008Q1 - 2011Q2 QLFS graduates to HEMIS data. Figures correspond to sample estimates and are unweighted.

B.1 Probabilistic linking and measurement error

To illustrate the implications of the measurement error that is likely to arise from the probabilistic linking approach described in Section 5.2, consider the case where one wishes to estimate the association between having attended an HDI and a given labour market outcome.[23] In its most basic form, the population relationship can be expressed as

$$y_{ic} = \alpha + \beta HDI_{ic} + \varepsilon_{ic} \tag{2}$$

where y_{ic} denotes the labour market outcome for individual i for whom the criterion variable $c_i^L = c$ (see equation (1)), β is the population parameter of interest, $HDI_{ic} \in \{0, 1\}$ is an indicator variable for whether individual i graduated from and HDI or not, and ε_{ic} is the additive error term.

HDI_{ic} is not observed in the LFS/QLFS data. However, under the assumption that the HEMIS data represents the population from which all graduates in the LFS/QLFS data are drawn, it is possible to estimate the population proportion of graduates from the group $c_i^H = c$ who graduated from HDIs as

$$\begin{aligned} \tilde{HDI}_c &= \frac{\sum_{j=1}^{N} \mathbf{1}\left(\mathbf{c}_j^H = c\right) \mathbf{1}\left(HEI_j = HDI\right)}{\sum_{k=1}^{N} 1\left(\mathbf{c}_k^H = c\right)} \\ &= E\left[HDI_i | C = c\right] \end{aligned} \tag{3}$$

such that

$$\tilde{HDI}_c = HDI_{ic} + u_{ic} \tag{4}$$

where u_{ic} captures the difference between the true HDI indicator variable and the imputed HDI proxy variable. Substituting (4) into (2) yields

$$\begin{aligned} y_{ic} &= \alpha + \beta\left(\tilde{HDI}_c - u_{ic}\right) + \varepsilon_{ic} \\ &= \alpha + \beta \tilde{HDI}_c + \left(\varepsilon_{ic} - \beta u_{ic}\right) \end{aligned}$$

By construction, $E\left(u\right) = 0$ and $\rho\left(\tilde{HDI}_c, u\right) = 0$. From this it follows that the OLS estimator of β will be consistent

$$\begin{aligned} \hat{\beta} &= \frac{cov\left(y_{ic}, \tilde{HDI}_c\right)}{var\left(\tilde{HDI}_c\right)} \\ &= \frac{cov\left(\beta \cdot HDI_{ic} + \varepsilon_{ic}, \tilde{HDI}_c\right)}{var\left(\tilde{HDI}_c\right)} \\ &= \frac{cov\left(\beta \cdot \left(\tilde{HDI}_c + u_{ic}\right) + \varepsilon_{ic}, \tilde{HDI}_c\right)}{var\left(\tilde{HDI}_c\right)} \\ \therefore p\lim \hat{\beta} &= \frac{\beta \sigma^2_{\tilde{HDI}_c}}{\sigma^2_{\tilde{HDI}_c}} = \beta \end{aligned}$$

[23] The discussion in this section is based on Pischke (2007).

While the parameter estimate, $\hat{\beta}_{H\tilde{D}I_c}$ will be consistent, the standard error, $\hat{se}_{H\tilde{D}I_c}$ will be inflated. This can be illustrated using Monte Carlo simulations on the HEMIS data.

Using the 2000 - 2013 aggregate HEMIS datasets, a hypothetical outcome variable, y, was defined as a function of the HDI indicator variable in the data

$$y_{ic} = \beta HDI_{ic} + \mu_{ic}$$

with $\beta = 0.1$ and $\mu_{ic} \sim N(0, 1)$. Next, $H\tilde{D}I_c$ was estimated via equation (1) in Section 5.2 for each of the three criteria variables. Finally, Monte Carlo simulations were used to calculate the average of $\hat{\beta}_{HDI}$, $\hat{\beta}_{H\tilde{D}I_c}$, $\hat{se}_{HDI}$, and $\hat{se}_{H\tilde{D}I_c}$ over 1000 trials for variously sized random samples. The results from these estimations are presented in Tables B.3 - B.5.

For each of the three probabilistic linking criteria used, it is clear that, on average, $\hat{\beta}_{H\tilde{D}I_c} \approx \hat{\beta}_{HDI} \approx \beta_{HDI}$ in sufficiently large samples. That is, $p\lim \hat{\beta}_{H\tilde{D}I_c} = p\lim \hat{\beta}_{HDI} = \beta$. However, regardless of the sample size, it remains the case that $\hat{se}_{H\tilde{D}I_c} > \hat{se}_{HDI}$. In other words, standard errors will be inflated whenever $H\tilde{D}I_c$ is used as a proxy for HDI_{ic}.

Table B.3: Average betas and standard errors obtained from Monte Carlo simulations using Criteria 1

Sample Size	$\hat{\beta}_{HDI}$	$\hat{se}_{HDI}$	$\hat{\beta}_{H\tilde{D}I}$	$\hat{se}_{H\tilde{D}I}$	$\hat{\beta}_{H\tilde{D}I}/\hat{\beta}_{HDI}$	$\hat{se}_{H\tilde{D}I}/\hat{se}_{HDI}$
100	0.100	0.210	0.091	0.496	0.906	2.366
200	0.104	0.148	0.098	0.350	0.937	2.368
500	0.099	0.093	0.098	0.220	0.991	2.357
1 000	0.098	0.066	0.110	0.156	1.117	2.366
2 000	0.099	0.047	0.097	0.110	0.981	2.356
5 000	0.099	0.030	0.098	0.070	0.989	2.359
10 000	0.099	0.021	0.098	0.049	0.993	2.358
20 000	0.099	0.015	0.098	0.035	0.990	2.358
100 000	0.099	0.007	0.099	0.016	0.999	2.358

NOTES: Figures represent the average betas and standard errors from the OLS estimations of (a) $y_{ic} = \alpha + \beta HDI_{ic} + \mu_{ic}$ and (b) $y_{ic} = \alpha + \beta H\tilde{D}I_c + \mu_{ic}$ using 1000 Monte Carlo trials for each of the variously sized random samples in the 2000 - 2013 aggregate HEMIS data. $\beta = 0.1$. Criterion 1 is given by the unique combinations of year of birth, race, gender, level of degree awarded (bachelor, postgraduate, etc), and SAQA field of study.

Table B.4: Average betas and standard errors obtained from Monte Carlo simulations using Criteria 2

Sample Size	$\hat{\beta}_{HDI}$	$\hat{se}_{HDI}$	$\hat{\beta}_{\tilde{HDI}}$	$\hat{se}_{\tilde{HDI}}$	$\hat{\beta}_{\tilde{HDI}}/\hat{\beta}_{HDI}$	$\hat{se}_{\tilde{HDI}}/\hat{se}_{HDI}$
100	0.093	0.210	0.086	0.562	0.931	2.682
200	0.096	0.148	0.100	0.397	1.040	2.687
500	0.098	0.093	0.103	0.250	1.042	2.677
1 000	0.098	0.066	0.115	0.177	1.167	2.689
2 000	0.099	0.047	0.103	0.125	1.043	2.677
5 000	0.101	0.030	0.099	0.079	0.972	2.681
10 000	0.100	0.021	0.102	0.056	1.017	2.681
20 000	0.100	0.015	0.101	0.040	1.011	2.681
100 000	0.099	0.007	0.101	0.018	1.021	2.681

NOTES: Figures represent the average betas and standard errors from the OLS estimations of (a) $y_{ic} = \alpha + \beta HDI_{ic} + \mu_{ic}$ and (b) $y_{ic} = \alpha + \beta \tilde{HDI}_c + \mu_{ic}$ using 1000 Monte Carlo trials for each of the variously sized random samples in the 2000 - 2013 aggregate HEMIS data. $\beta = 0.1$. Criterion 2 is given by the unique combinations of year of birth, race, gender, and level of degree awarded (bachelor, postgraduate, etc).

Table B.5: Average betas and standard errors obtained from Monte Carlo simulations using Criteria 3

Sample Size	$\hat{\beta}_{HDI}$	$\hat{se}_{HDI}$	$\hat{\beta}_{\tilde{HDI}}$	$\hat{se}_{\tilde{HDI}}$	$\hat{\beta}_{\tilde{HDI}}/\hat{\beta}_{HDI}$	$\hat{se}_{\tilde{HDI}}/\hat{se}_{HDI}$
100	0.094	0.210	0.089	0.583	0.942	2.780
200	0.099	0.148	0.105	0.412	1.059	2.793
500	0.101	0.094	0.099	0.259	0.980	2.771
1 000	0.095	0.066	0.115	0.184	1.215	2.786
2 000	0.102	0.047	0.103	0.130	1.011	2.774
5 000	0.099	0.030	0.099	0.082	0.998	2.778
10 000	0.100	0.021	0.101	0.058	1.009	2.778
20 000	0.100	0.015	0.101	0.041	1.008	2.778
100 000	0.099	0.007	0.102	0.018	1.027	2.778

NOTES: Figures represent the average betas and standard errors from the OLS estimations of (a) $y_{ic} = \alpha + \beta HDI_{ic} + \mu_{ic}$ and (b) $y_{ic} = \alpha + \beta \tilde{HDI}_c + \mu_{ic}$ using 1000 Monte Carlo trials for each of the variously sized random samples in the 2000 - 2013 aggregate HEMIS data. $\beta = 0.1$. Criterion 3 is given by the unique combinations of year of birth, race, and gender.

B.2 Composition of graduates in HEMIS and LFS/QLFS data

Table B.6: Proportion of graduates from the 2000 - 2013 HEMIS data sample, by HEI type

	Tech	Comp	Trad	UTech	HDI	Clust1	Clust2	Clust3
				Gender				
Male	3.0	20.9	68.1	8.0	32.4	47.5	42.3	10.2
Female	1.9	23.5	67.4	7.2	34.7	46.4	44.0	9.6
				Race				
Black	3.1	27.3	58.8	10.9	48.0	32.9	46.2	20.9
Coloured	2.8	16.2	68.9	12.0	41.6	39.3	60.4	0.3
Indian	1.8	21.1	72.4	4.7	40.2	60.9	38.1	1.0
White	1.7	18.3	76.4	3.6	15.1	61.5	38.1	0.4
				Birth Cohort				
1930s	4.5	17.6	76.7	1.2	13.4	43.3	53.2	3.5
1940s	5.3	14.8	77.2	2.7	23.9	41.4	53.2	5.4
1950s	5.3	18.2	71.4	5.1	30.2	40.7	52.2	7.1
1960s	5.6	22.7	65.3	6.4	29.8	38.9	53.4	7.7
1970s	4.8	20.7	67.4	7.1	29.7	34.6	55.8	9.6
1980s	0.5	23.9	67.1	8.6	36.0	50.0	39.5	10.5
1990s	0.0	20.4	73.9	5.7	40.5	59.3	31.2	9.5
All	2.4	22.4	67.7	7.5	33.7	46.8	43.3	9.8

NOTES: Figures represent the actual proportions of graduates in HEMIS administrative data sample who graduated from specific types of HEIs over the period 2000 - 2013. Tech: Technikon; Comp: Comprehensive university; Trad: Traditional university; UTech: University of Technology; HDI: Historically Disadvantaged Institution; Clust1: CHET Cluster 1; Clust2: CHET Cluster 2; Clust3: CHET Cluster 3.

Table B.7: Estimated proportion of graduates in the 2000 - 2015 LFS/QLFS sample, by HEI type

	Tech	Comp	Trad	UTech	HDI	Clust1	Clust2	Clust3
				Gender				
Male	14.0	13.5	65.9	6.6	25.2	34.7	57.5	7.8
Female	8.8	17.2	68.0	6.1	29.0	32.7	58.9	8.4
				Race				
Black	12.2	16.1	63.5	8.2	41.5	28.3	55.6	16.1
Coloured	11.5	12.8	64.9	10.8	39.2	30.8	68.8	0.4
Indian	8.7	16.0	69.9	5.3	30.0	42.9	55.5	1.6
White	10.7	14.9	70.6	3.7	9.1	38.8	59.6	1.6
				Birth Cohort				
1930s	16.5	1.1	82.4	0.0	3.3	23.1	76.9	0.0
1940s	12.1	9.0	77.2	1.7	14.3	34.9	61.3	3.8
1950s	15.5	13.0	66.4	5.2	21.7	30.1	64.1	5.8
1960s	14.9	15.8	62.6	6.6	24.1	28.8	64.4	6.8
1970s	10.1	15.0	68.4	6.5	31.2	30.8	59.7	9.5
1980s	0.9	21.0	68.9	9.2	38.4	48.6	39.7	11.7
1990s	0.0	21.1	72.1	6.7	44.9	56.8	32.8	10.4
All	11.3	15.4	67.0	6.3	27.1	33.7	58.2	8.1

NOTES: Figures represent the proportions of working-age graduates in the labour force survey data sample who are estimated to have graduated from specific types of HEI over the period 2000 - 2013, based on probabilistic linking between aggregate HEMIS data and LFS data. Tech: Technikon; Comp: Comprehensive university; Trad: Traditional university; UTech: University of Technology; HDI: Historically Disadvantaged Institution; Clust1: CHET Cluster 1; Clust2: CHET Cluster 2; Clust3: CHET Cluster 3. Estimates are unweighted.

Table B.8: Estimated proportion of graduates in the working-age population, by HEI type

	Tech	Comp	Trad	UTech	HDI	Clust1	Clust2	Clust3
				Gender				
Male	12.2	14.5	66.6	6.7	24.5	36.5	56.0	7.6
Female	7.5	18.5	68.0	6.0	27.9	34.2	57.8	8.1
				Race				
Black	10.4	17.9	62.8	8.9	42.6	29.4	54.2	16.4
Coloured	10.8	13.2	65.1	10.9	39.0	31.0	68.6	0.4
Indian	7.7	17.7	69.2	5.5	31.0	43.4	54.9	1.7
White	9.8	15.3	71.2	3.7	9.4	40.2	58.2	1.6
				Birth Cohort				
1930s	16.3	1.2	82.5	0.0	2.3	18.0	82.0	0.0
1940s	12.2	10.7	75.5	1.6	11.6	33.6	63.0	3.4
1950s	13.9	14.6	66.4	5.2	19.6	31.9	62.9	5.2
1960s	13.4	16.7	63.4	6.5	22.1	30.2	63.6	6.2
1970s	9.3	15.6	68.5	6.6	30.2	31.8	58.8	9.4
1980s	0.8	21.2	69.2	8.8	36.9	50.1	38.9	11.0
1990s	0.0	20.5	73.1	6.4	42.3	59.3	31.3	9.3
All	9.9	16.4	67.3	6.4	26.1	35.4	56.8	7.8

NOTES: Figures represent the proportions of working-age graduates in the South African population who are estimated to have graduated from specific types of HEI over the period 2000 - 2013, based on probabilistic linking between aggregate HEMIS data and LFS/QLFS data. Tech: Technikon; Comp: Comprehensive university; Trad: Traditional university; UTech: University of Technology; HDI: Historically Disadvantaged Institution; Clust1: CHET Cluster 1; Clust2: CHET Cluster 2; Clust3: CHET Cluster 3. Estimates are weighted.

Appendix C Graduate unemployment/employment probit tables

C.1 Probability of narrow unemployment

Table C.1: Estimated probability of narrow unemployment for graduates, by HEI type

	(1)	(2)	(3)
Age	−0.200***	−0.196***	−0.199***
Age^2	0.002***	0.002***	0.002***
Coloured	−0.494***	−0.444***	−0.528***
Indian	−0.313***	−0.328***	−0.614***
White	−0.713***	−0.719***	−0.797***
Female	−0.006	−0.022	−0.020
Honnours	−0.137***	−0.185***	−0.189***
Masters+	−0.110**	−0.098**	−0.106**
Technikon		−0.153*	−0.200**
Technology		−0.395**	−0.650***
Comprehensive		0.669***	0.451***
Technikon × Coloured			−0.314
Technikon × Indian			0.467*
Technikon × White			0.065
Technology × Coloured			0.659*
Technology × Indian			1.210***
Technology × White			0.848**
Comprehensive × Coloured			0.223
Comprehensive × Indian			0.873***
Comprehensive × White			0.146
Observations	66 167	66 139	66 139
P-value	0.000	0.000	0.000
Area under ROC curve	0.794	0.796	0.796
Sensitivity	73.028	72.696	72.523
Specificity	72.807	73.255	73.525
Cutoff used	0.05	0.05	0.05

NOTES: *Significant at the 10% level **Significant at the 5% level *** Significant at the 1% level. Significance levels are based on linearised robust standard errors which have been adjusted for complex survey design. Estimates are weighted and were estimated using probit regression. The dependent variable is *narrow unemployment*. The sample includes only working-age graduates in the narrow labour force from the pooled 2000a LFS - 2015QLFSQ2 data. Regressions include controls for the survey period, province, and enrolment at an educational institution (suppressed in output). Reference categories are as follows: Race (Black); Gender (Male); Qualification (Bachelor's degree or equivalent); HEI (Traditional University). The chosen cut-off value for the calculated prediction sensitivity and specificity is equal to the proportion of the estimation sample who are narrowly unemployed.

Table C.2: Estimated probability of narrow unemployment for graduates, by HEI historical status

	(1)	(2)	(3)
Age	−0.200***	−0.197***	−0.199***
Age^2	0.002***	0.002***	0.002***
Coloured	−0.494***	−0.483***	−0.505***
Indian	−0.313***	−0.285***	−0.002
White	−0.713***	−0.640***	−0.546***
Female	−0.006	−0.009	−0.012
Honnours	−0.137***	−0.115***	−0.105***
Masters+	−0.110**	−0.092**	−0.069
HDI		0.219***	0.375***
HDI × Coloured			0.071
HDI × Indian			−0.806***
HDI × White			−0.381
Observations	66 167	66 139	66 139
P-value	0.000	0.000	0.000
Area under ROC curve	0.794	0.794	0.795
Sensitivity	73.028	72.437	72.206
Specificity	72.807	73.133	73.517
Cutoff used	0.05	0.05	0.05

NOTES: *Significant at the 10% level **Significant at the 5% level *** Significant at the 1% level. Significance levels are based on linearised robust standard errors which have been adjusted for complex survey design. Estimates are weighted and were estimated using probit regression. The dependent variable is *narrow unemployment*. The sample includes only working-age graduates in the narrow labour force from the pooled 2000a LFS - 2015QLFSQ2 data. Regressions include controls for the survey period, province, and enrolment at an educational institution (suppressed in output). Reference categories are as follows: Race (Black); Gender (Male); Qualification (Bachelor's degree or equivalent); HEI historical status (Historically Advantaged Institution (HAI)). The chosen cut-off value for the calculated prediction sensitivity and specificity is equal to the proportion of the estimation sample who are narrowly unemployed.

Table C.3: Estimated probability of narrow unemployment for graduates, by HEI cluster

	(1)	(2)	(3)
Age	−0.200***	−0.206***	−0.203***
Age^2	0.002***	0.002***	0.002***
Coloured	−0.461***	−0.380***	−7.935*
Indian	−0.276***	−0.180***	−5.772***
White	−0.671***	−0.578***	−0.535
Female	−0.027	−0.028	−0.027
Honours	−0.139***	−0.068*	−0.123***
Masters+	−0.120***	−0.027	−0.080
Cluster 1 HEI		−0.593***	−0.345**
Cluster 2 HEI		−0.405***	−0.503***
Cluster 1 × Coloured			7.500
Cluster 1 × Indian			5.174**
Cluster 1 × White			−0.298
Cluster 2 × Coloured			7.632
Cluster 2 × Indian			5.990***
Cluster 2 × White			0.114
Observations	55 018	54 991	54 991
P-value	0.000	0.000	0.000
Area under ROC curve	0.786	0.786	0.788
Sensitivity	71.064	70.301	69.248
Specificity	73.232	74.101	75.091
Cutoff used	0.05	0.05	0.05

NOTES: *Significant at the 10% level **Significant at the 5% level *** Significant at the 1% level. Significance levels are based on linearised robust standard errors which have been adjusted for complex survey design. Estimates are weighted and were estimated using probit regression. The dependent variable is *narrow unemployment*. The sample includes only working-age graduates in the narrow labour force from the pooled 2004a LFS - 2015QLFSQ2 data. Regressions include controls for the survey period, province, and enrolment at an educational institution (suppressed in output). Reference categories are as follows: Race (Black); Gender (Male); Qualification (Bachelor's degree or equivalent); HEI cluster (Cluster 3 Institution). The chosen cut-off value for the calculated prediction sensitivity and specificity is equal to the proportion of the estimation sample who are narrowly unemployed.

C.2 Probability of employment

Table C.4: Estimated probability of employment for graduates, by HEI type

	(1)	(2)	(3)
Age	0.264***	0.264***	0.264***
Age^2	−0.003***	−0.003***	−0.003***
Coloured	0.259***	0.218***	0.215***
Indian	−0.113***	−0.091***	−0.031
White	0.063***	0.081***	0.093***
Female	−0.404***	−0.376***	−0.378***
Honnours	0.148***	0.181***	0.185***
Masters+	0.246***	0.214***	0.218***
Technikon		0.175**	0.281***
Technology		0.380***	0.405**
Comprehensive		−0.662***	−0.639***
Technikon × Coloured			−0.105
Technikon × Indian			−0.124
Technikon × White			−0.186
Technology × Coloured			−0.015
Technology × Indian			−0.247
Technology × White			0.029
Comprehensive × Coloured			0.117
Comprehensive × Indian			−0.190
Comprehensive × White			0.030
Observations	73 808	73 753	73 753
P-value	0.000	0.000	0.000
Area under ROC curve	0.745	0.748	0.748
Sensitivity	71.791	71.835	71.767
Specificity	63.248	63.260	63.332
Cutoff used	0.85	0.85	0.85

NOTES: *Significant at the 10% level **Significant at the 5% level *** Significant at the 1% level. Significance levels are based on linearised robust standard errors which have been adjusted for complex survey design. Estimates are weighted and were estimated using probit regression. The dependent variable is *employment*. The sample includes only working-age graduates from the pooled 2000a LFS - 2015QLFSQ2 data. Regressions include controls for the survey period, province, and enrolment at an educational institution (suppressed in output). Reference categories are as follows: Race (Black); Gender (Male); Qualification (Bachelor's degree or equivalent); HEI (Traditional University). The chosen cut-off value for the calculated prediction sensitivity and specificity is equal to the proportion of the estimation sample who are employed.

Table C.5: Estimated probability of employment for graduates, by HEI historical status

	(1)	(2)	(3)
Age	0.264***	0.263***	0.264***
Age^2	−0.003***	−0.003***	−0.003***
Coloured	0.259***	0.257***	0.183**
Indian	−0.113***	−0.121***	−0.375***
White	0.063***	0.033	−0.174***
Female	−0.404***	−0.404***	−0.399***
Honnours	0.148***	0.142***	0.128***
Masters+	0.246***	0.244***	0.212***
HDI		−0.087	−0.421***
HDI × Coloured			0.141
HDI × Indian			0.651***
HDI × White			0.909***
Observations	73 808	73 753	73 753
P-value	0.000	0.000	0.000
Area under ROC curve	0.745	0.745	0.746
Sensitivity	71.791	71.701	71.540
Specificity	63.248	63.179	63.513
Cutoff used	0.85	0.85	0.85

NOTES: *Significant at the 10% level **Significant at the 5% level *** Significant at the 1% level. Significance levels are based on linearised robust standard errors which have been adjusted for complex survey design. Estimates are weighted and were estimated using probit regression. The dependent variable is *employment*. The sample includes only working-age graduates from the pooled 2000a LFS - 2015QLFSQ2 data. Regressions include controls for the survey period, province, and enrolment at an educational institution (suppressed in output). Reference categories are as follows: Race (Black); Gender (Male); Qualification (Bachelor's degree or equivalent); HEI historical status (Historically Advantaged Institution (HAI)). The chosen cut-off value for the calculated prediction sensitivity and specificity is equal to the proportion of the estimation sample who are employed.

Table C.6: Estimated probability of employment for graduates, by HEI cluster

	(1)	(2)	(3)
Age	0.260***	0.274***	0.271***
Age^2	−0.003***	−0.003***	−0.003***
Coloured	0.227***	0.137***	0.013
Indian	−0.153***	−0.278***	3.365***
White	0.027	−0.083***	0.293
Female	−0.391***	−0.390***	−0.387***
Honours	0.151***	0.045	0.083***
Masters+	0.253***	0.100**	0.125***
Cluster 1 HEI		0.819***	0.579***
Cluster 2 HEI		0.489***	0.646***
Cluster 1 × Coloured			0.224
Cluster 1 × Indian			−3.441***
Cluster 1 × White			−0.063
Cluster 2 × Coloured			0.048
Cluster 2 × Indian			−3.841***
Cluster 2 × White			−0.574
Observations	61 190	61 146	61 146
P-value	0.000	0.000	0.000
Area under ROC curve	0.748	0.749	0.751
Sensitivity	72.298	72.150	72.049
Specificity	62.991	63.539	63.880
Cutoff used	0.86	0.86	0.86

NOTES: *Significant at the 10% level **Significant at the 5% level *** Significant at the 1% level. Significance levels are based on linearised robust standard errors which have been adjusted for complex survey design. Estimates are weighted and were estimated using probit regression. The dependent variable is *employment*. The sample includes only working-age graduates from the pooled 2004a LFS - 2015QLFSQ2 data. Regressions include controls for the survey period, province, and enrolment at an educational institution (suppressed in output). Reference categories are as follows: Race (Black); Gender (Male); Qualification (Bachelor's degree or equivalent); HEI Cluster (Cluster 3 Institution). The chosen cut-off value for the calculated prediction sensitivity and specificity is equal to the proportion of the estimation sample who are employed.

C.3 Predicted average marginal racial differentials in graduate narrow unemployment and employment

Table C.7: Predicted average racial differentials (percentage point difference relative to Blacks) in narrow unemployment and employment rates (%) for graduates by race and HEI type (2000 - 2015)

	Narrow Unemployment			Employment		
HEI Type	**Coloured**	**Indian**	**White**	**Coloured**	**Indian**	**White**
Traditional	−5.1***	−5.6***	−6.4***	3.9***	−0.6	1.8***
Technikon	−4.8***	−1.4	−4.5***	1.5	−2.5	−1.5
Technology	0.7	5.1	0.3	2.3	−4.2	1.5
Comprehensive	−5.9	6.5	−10.5***	9.4	−7.1	3.7
HDI	−5.9***	−8.8***	−9.4***	7.2***	6.2***	13.6***
HAI	−3.8***	−0.0	−4.0***	2.9**	−7.8**	−3.3***
Cluster 1	−4.3***	−5.3***	−6.3***	3.9**	−1.5**	3.9***
Cluster 2	−2.6***	2.7*	−3.3***	1.0	−10.4	−5.7***
Cluster 3	−13.4***	−13.4***	−7.9	0.4	27.8	7.8

NOTES: [a]Figures reflect the estimated average percentage point difference in the predicted graduate narrow unemployment rate for the respective race groups relative to Black graduates by HEI type over the period 2000 - 2015 and are based on the marginal predictions from the regressions in columns (3) of Tables C.1 - C.3. [b]Figures reflect the estimated average percentage point difference in the predicted graduate employment rates for the respective race groups relative to Black graduates by HEI type over the period 2000 - 2015 and are based on the marginal predictions from the regressions in columns (3) of Tables C.1 - C.3. Predictions generated by setting the relevant HEI type proxy variables equal to 1 or 0. E.g. the predicted rates for graduates from traditional HEIs was generated using Technikon = 0, Technology = 0, and Comprehensive = 0, whereas the predicted rates for graduates from Cluster 2 HEIs was generated using Cluster 1 = 0 and Cluster 2 = 1.All other variables were kept at their observed values in the data when calculating the respective expected graduate unemployment/employment rates. *Significant at the 10% level **Significant at the 5% level *** Significant at the 1% level. Significance levels are based on linearised robust standard errors which have been adjusted for complex survey design. Estimates are weighted.

Table C.8: Predicted average racial differentials (% difference relative to Blacks) in narrow unemployment and employment rates (%) for graduates by race and HEI type (2000 - 2015)

	Narrow Unemployment			Employment		
HEI Type	**Coloured**	**Indian**	**White**	**Coloured**	**Indian**	**White**
Traditional	−63.0***	−69.1***	−79.0***	4.6***	−0.7	2.1***
Technikon	−84.2***	−24.6	−78.9***	1.7	−2.8	−1.7
Technology	29.2	212.5	12.5	2.5	−4.6	1.6
Comprehensive	−36.4	40.1	−64.8***	13.7	−10.3	5.4
HDI	−51.8***	−77.2***	−82.5***	9.1***	7.9***	17.3***
HAI	−63.3***	−0.0	−66.7***	3.3**	−8.9**	−3.8***
Cluster 1	−55.8***	−68.8***	−81.8***	4.5**	−1.7**	4.5***
Cluster 2	−44.1***	45.8*	−55.9***	1.1	−11.9	−6.5***
Cluster 3	= 100.0***	= 100.0***	−59.0	0.6	38.6	10.8

NOTES: [a]Figures reflect the estimated average percentage (%) difference in the predicted graduate narrow unemployment rate for the respective race groups relative to Black graduates by HEI type over the period 2000 - 2015 and are based on the marginal predictions from the regressions in columns (3) of Tables C.1 - C.3. [b]Figures reflect the estimated average percentage (%) difference in the predicted graduate employment rates for the respective race groups relative to Black graduates by HEI type over the period 2000 - 2015 and are based on the marginal predictions from the regressions in columns (3) of Tables C.1 - C.3. Predictions generated by setting the relevant HEI type proxy variables equal to 1 or 0. E.g. the predicted rates for graduates from traditional HEIs was generated using Technikon = 0, Technology = 0, and Comprehensive = 0, whereas the predicted rates for graduates from Cluster 2 HEIs was generated using Cluster 1 = 0 and Cluster 2 = 1.All other variables were kept at their observed values in the data when calculating the respective expected graduate unemployment/employment rates. *Significant at the 10% level **Significant at the 5% level *** Significant at the 1% level. Significance levels are based on linearised robust standard errors which have been adjusted for complex survey design. Estimates are weighted.