University of Cape Town

Is Transformation Surviving?

Author:
Michael DALY
dlymic006@myuct.ac.za

Supervisor:
Dr. Corné VAN WALBEEK
cwalbeek@gmail.com

May 2016

Abstract:

Recent political events in South Africa have emphasized the importance of faculty diversity. Very little research has considered why it is the case that 20 years after the end of Apartheid, only 14% of professors are black. Or, why the University of Cape Town does not have a single black South African woman who is a full professor. Is it the case that black faculty are discriminated against during the hiring process or is it the case that black faculty depart at significantly higher rates than white faculty as sometimes suggested? Further, how do race, education and institutional factors interact in determining diversity levels? Survival analysis methods coupled with a novel data-set consisting of detailed administrative employee records, proxies for performance and various socio-economic variables are employed to test various hypotheses related to these questions. The findings suggest that the level of diversity is of secondary importance to an employee's race when determining the likelihood of survival at the University of Cape Town. In addition, findings suggest that the university is performing far better in terms of gender equality than racial equality. Overall, race matters, not only through a direct correlation between employee race and tenure length, but also through indirect effects where employees who differ significantly from others in their respective faculty (in terms of race, tenure, age, education and gender) face increased rates of departure from the work place.

Acknowledgements

Short and sweet

I would like to thank everyone that has helped me throughout this journey.

Without you, it would not have been possible.

And, I would especially like to thank my supervisor

Dr. Corne Van Walbeek

Contents

Introduction

Employee turnover has significant cost implications for most organizations, particularly where the employees involved poses specialized skills or require significant training. As a result, a detailed understanding of the turnover rates, and reasons for the turnover are of importance when developing strategies to increase retention. In recent years, the University of Cape Town has faced increased criticism both from within the organization and from individuals and groups outside of the organization regarding the low number of black faculty it employs. These discussions are situated within the context of a broader debate surrounding transformation in higher education in South Africa. On the one hand the university would want to reduce turnover due to the significant costs involved (recruitment, lost productivity, set-up costs and training), but on the other hand the university desires to increase the level of diversity among its academic staff necessitating a certain level of turnover in the absence of expansion. Thus, the reduction of overall employee turnover needs to be balanced against the goal of increasing recruitment and retention of faculty of colour in particular.

Using a novel data set I examine in detail the dynamic nature of turnover at the University of Cape Town. In particular I examine (1) how individual demographic and non-demographic traits influence length of tenure, and (2) what happens when individuals who are different from one another work together. For example, how different is the experience of being male in a female-dominated department to being female in a male-dominated department? By doing so, I hope to provide information useful to policy makers that extends beyond the primarily political discussion that has occurred in the public media in South Africa. An understanding of the various factors that drive differences in tenure length and influence the likelihood of premature departure is an important aid for

administrators and other stakeholders who need to make hiring and organizational policy decisions. These are particularly interesting questions given the changes that have taken place in the South African labour market over the last two decades. Prior to 1994, black Africans in particular had very little access to the academic employment market while today, at least in theory, different races have equal access.

Job tenure occupies a unique position in the economics literature. In part, this can be attributed to the fact that the study of job tenure poses unique challenges given the nature of the data and the underlying process being studied. In fact, prior to the development of statistical methods such as survival analysis, economists were restricted in their ability to study employment tenure in an empirically rigorous manner given the frequent presence of censored[1] data, which render traditional regression methods invalid.

A unique advantage of the data set I construct for this paper is its longitudinal nature, length of coverage and level of detail. The structure of the data set allows for the use of survival analysis methods along with detailed trend analysis. Survival analysis methods are able to better characterize dynamic changes in tenure and employment than typical cross-sectional methods, which provide a point-in-time analysis. Given the nature of university employment – typically characterized by relatively low levels of turnover and long service, a

1 In the context of employment data, censoring most typically occurs when an employee remains in employment at the end of the study period. The observation is said to be censored as we know that the employee will at some point in the future experience a termination, but not the exact time at which this termination will occur.

dataset spanning a significant number of years was necessary to capture the desired level of analytic detail[2].

The first section in this paper is structured so as to provide an understanding of the existing literature and an overview of the situation at the University of Cape Town, followed by a detailed examination of UCT employee statistics. A theoretical model regarding the processes driving employee turnover and motivation for the use of survival analysis methods proceeds next. The third section consists of detailed discussion on variable construction, models employed (both semi-parametric and non-parametric) and the empirical results. The paper ends with a discussion of the results, notes limitations to the study and makes recommendations for areas of future investigation.

An overview of the related literature

Turnover has been studied extensively, particularly within the field of organizational psychology. In general, studies suggest that external factors, work-related factors and personal characteristics are strongly related to employee turnover[3]. Within the field of economics, the study of academic tenure can be related to two distinct bodies of literature. Firstly, it relates to the literature regarding workplace turnover and the determinants thereof. Secondly, it can be related to the study of workplace diversity and the consequences of interaction among workers who differ from one another. This literature review

2 While SA specific data on long-term employment tenure was not available, the US Bureau of Labor Statistics finds that wage and salary workers in the US public Sector had nearly double the median tenure of private sector employees. Further, they find that employees in "education, training and library" had the third highest tenure among all employees (Bureau of Labor Statistics, 2014).

3 Meta-analyses covering these studies have been produced, Cotton and Tuttle cover in excess of 120 of these studies and a more recent meta-analysis by Griffeth, Hom and Gaertner examines specific determinants in more detail (Griffeth, Hom, & Gaertner, 2000) (Cotton & Tuttle, 1986).

will focus on prior work in the economics literature most relevant to the current study.

2.1 Origins of job tenure in economics

Early economics literature on job tenure focused primarily on the theoretical issues related to the employer – employee job match. The theoretical focus was in part driven by the absence of suitable empirical data with which to test economic theories in addition to the limitation imposed by available computing power. Early scholarship suggested that the quality of the employer – employee match could only be determined subsequent to the match being made and the worker having actually worked (Jovanovic, 1979). In particular, Jovanovic's model established a link between workers remaining in jobs where their productivity was relatively high, and leaving those where their productivity was relatively low, thereby suggesting that mismatch between employer and employee is likely to be detected sooner rather than later (Jovanovic, 1979). These early studies differ in nature to this paper as they did not focus on specific determinants that drove the quality of employer employee matches.

With the increased availability of detailed workplace- and employee-level data, empirical studies on job tenure have become prevalent. Such studies have identified a broad range of both work-place and individual-level characteristics as being related to the likelihood that a worker will remain in employment at a given firm. These studies have used a wide variety of estimation strategies and data sources ranging from firm level data to national household surveys. Linked work-place and individual-level panel and cross-sectional data studies have found that a range of demographic (age, gender and race), job-related, educational and occupational variables are important in explaining job tenure (Mumford & Smith, 2004). A number of studies have considered the effect of

gender on turnover and find that females generally experience higher levels of turnover, especially if combined with lower levels of education (Royalty, 1998). Additionally, studies suggest gender differences between job-to-job turnover and job-to-unemployment turnover, with females more inclined to job-to-unemployment turnover than males (Sousa-Poza & Sousa-Poza, 2007) (Royalty, 1998).

A distinct body of literature has focused on the relationship between job satisfaction and turnover where findings typically suggest that higher levels of job satisfaction are associated with lower levels of employee turnover. A contemporary study using Swiss household survey data and an ordinal 10-point scale to measure job satisfaction supports this relationship between job satisfaction and turnover (Sousa-Poza & Sousa-Poza, 2007). Economists have also studied the role of education, turnover and job satisfaction among highly educated academics. Evidence from a sample of doctoral professionals in science, technology, engineering and mathematics (STEM) fields suggests that the process of job satisfaction formation differs between men and women (Bender & Heywood, 2006). Additionally, job satisfaction, job commitment and quality of work life are related to job tenure in the small and medium enterprise sector in Zimbabwe (Chinomona & Dhurup, 2014).

Following the evidence that various measurable characteristics are correlated to labour market turnover, some studies have gone on to examine the effectiveness of various programs at increasing employee retention. There is some evidence to suggest that relatively simple training programs can increase early retention among professional employees (Mattox & Jinkerson, 2005). Other research suggests that increased support programs for new faculty are of importance (Piercy, et al., 2005). In addition, the need for careful monitoring of progress

made towards increasing the level of faculty diversity at higher educational institutions has been suggested as important (Weinberg, 2008). I do not discuss these programs in detail as they are not the focus of this thesis. A study of faculty turnover at the University of Wisconsin Madison suggests that the reason faculty left fell into three distinct categories: involuntary termination (retrenchment, dismissal), voluntary termination (dissatisfied with position, found better opportunities or a higher salary elsewhere), or the end of career termination (retirement or death) (Harrigan, 1999).

2.2 Effects of workplace diversity

A distinct, but related area of study suggests that workplace turnover is a consequence of workplace diversity levels rather than particular individual-level characteristics. That is, the focus falls on the level of dissimilarity between some focal worker and the average worker within a particular organizational unit or department. This implies that high rates of turnover among specific demographic groups are not dependent on specific demographic traits, but rather the result of interactions between different groups of employees and that those who share characteristics with their fellow employees experience higher levels of job satisfaction. Lazear (1999) suggested that a firm faces additional challenges when workers of different cultures and national origins are employed and work together (Lazear, 1999). There are both costs and benefits to a multi-cultural workforce. Integrating different laws, languages, cultures and customs in a given firm can facilitate division of labour and mutual learning within organizational units (Lazear, 1999). On the other hand increased communication costs could outweigh the expected returns of worker dissimilarity (Lazear, 1999). Studies that have examined diversity in the work place generally support

self-categorization theory[4] and suggest that individuals are motivated to maintain their group identity and seek to maximize intergroup distinctiveness (Tsui, Egan, & O'Reilly, 1992).

Two contradictory theoretic models undergird the possible outcomes associated with workplace diversity. *Social contact theory* suggests that as diversity increases individuals interact with members of other groups more frequently, thereby disproving group stereotypes (Tolbert, Simons, Andrews, & Rhee, 1995). On the other hand, *competition theory* suggests that members of socially defined groups compete for the control of scarce resources. Consequently, increases in the proportionate size of a minority group lead to heightened conflict between groups with the result that the minority group faces increased rates of turnover (Tolbert, Simons, Andrews, & Rhee, 1995). Empirical evidence suggests that conflict theory may prevail in male-dominated departments – at least until some given threshold of female faculty is reached (Tolbert, Simons, Andrews, & Rhee, 1995).

Expanding upon the idea that the level of diversity in the workplace matters, Kurtulus (2011) examined the effects of individual worker dissimilarity in terms of demographic traits (age, gender, race and education) and simultaneously in terms of work function, firm tenure and wages on performance of workers within the same division, using data from a large US health services firm (Kurtulus, 2011). That particular study found a significant negative relationship between racial dissimilarity and worker performance in the cross-sectional setting (Kurtulus, 2011). However, this result disappeared in a panel differenced model suggesting that unobserved employee characteristics are important in determining the relationship between dissimilarity and survival. (Kurtulus, 2011).

4 Self-categorization of individuals into psychological groups

This study suggested that both age differences and differences in worker tenure affect the performance of the focal worker (Kurtulus, 2011).

2.2 Diversity of faculty

The issue of diversity in academia is by no means unique to South Africa. A large number of leading universities in the United States have undertaken measures, with varying levels of, success to increase the representation of minority faculty (Wood, 2008). In addition, the National Bureau of Economic Research has long been collecting academic employee data for academics at American colleges and universities for research purposes (Ehrenberg, Kasper, & Rees, 1990). Analysis of such data has revealed high levels of stability in turnover rates and evidence to suggest that salary is positively related to retention (Ehrenberg, Kasper, & Rees, 1990).

2.3 Increasing diversity levels

Extending the discussion on faculty diversity, evidence suggests that women and members of under-represented minority groups are judged more fairly in the hiring process when they make up at least thirty percent of the applicant pool (Sackett, DuBois, & Noe, 1991). It is important to note that this level of participation is difficult to achieve, particularly in STEM fields where relatively few women obtain doctoral degrees. Nevertheless, the influence of early participation was recently confirmed in a study that suggested that increasing the level of diversity in the early stages of the applicant pool could improve hiring outcomes with respect to diversity (Bilmoria & Buch, 2010).

In summary, the literature regarding job turnover has undergone an evolutionary process over the last two decades, partly driven by the increased availability of relevant data and advances in statistical methodology. Studies have varied in their focus, with some aiming to identify characteristics that explain job tenure.

Here, numerous characteristics, both at the individual and the workplace levels, have shown to be significant. Other studies have emphasised that it is not one particular characteristic that influences workplace tenure decisions and match, but rather an individual's demographic characteristics in relation to fellow employees within a department or division. That is, tenure may be the product of workplace diversity. Moreover, the literature regarding faculty diversity might seem to be a new theme in the South African context, but it has long been considered by leading universities in the United States, who view enhancing diversity has an important institutional goal. Though the advantages of increased faculty diversity continue to stay primarily in the realm of conjecture, the subsequent suggestion that increased faculty diversity promotes intercultural learning, as well as new ideas and potential for improved mentorship relationships, remain as desirable outcomes.

2.4 The debate at the University of Cape Town

The University of Cape Town has been the center of media attention in recent months. In much of the media debate, the university has faced heavy criticism and accusations of what some claim is its apparent hostility towards transformation (Groundup Staff, 2015), (Nogcinisa, 2015) and (O'Connell, 2015). The transformation debate at the university has focused on three distinct aspects vital to the university community as a whole: (1) transformation of the student body, (2) transformation of the curriculum and (3) transformation of the faculty. While all of these issues are interesting in their own right, this paper will focus exclusively on the third. While not a unique problem faced by the University of Cape Town, it has borne the brunt of criticism regarding the apparent lack of minority representation in the academic staff bodies of South Africa's leading universities. While much of the debate has focused on race, the

issue of gender representation is equally important given the fact that academia, both globally and in South Africa, has traditionally been male dominated.

In a graduation ceremony that took place in December 2014, Graca Machel made the comment that who is being taught and how they are being taught is more important than who is doing the teaching (University of Cape Town, 2014). This particular statement has drawn criticism from a number of different commentators who disagree with varying degrees to the statement, including a strongly worded letter penned by a group of academics who self-identify as black (Haupt, 2015), (Kessi, et al., 2015). These critics have gone as far as to suggest that the underrepresentation of black academics is the result of discriminatory hiring practices at the university (Cape Times, 2015).

On the other hand, the university's vice chancellor, Dr Max Price, suggests that the reason for limited success in academic staff transformation initiatives has been the difficulty in attracting suitably qualified minority applicants (Price, 2014). Moreover, the length of time required to achieve the rank of a full professor is seen as another constraint to achieving the university's transformation objectives (Price, 2014). A better understanding of these issues will hopefully assist in bridging the differences between administrators who view the shortage of qualified minority academics as the major transformation hurdle and black academics within the university who argue that the primary cause is in fact a racially prejudiced institutional culture. Surprisingly, the university's 2015-2020 employment equity plan does not identify retention of "minority" employees as a key issue but does note five key barriers to achieving employment equity targets: 1. a limited pool of black academics, 2. high retention rates among all employees restricting change in the racial profile of employees, 3. The difficulty of recruiting graduates into doctoral studies and

then academic careers, 4. a perception that the university does not have a culture of inclusivity and 5. a perception among black and women staff that the university does not do enough to develop their careers (University of Cape Town, 2015).

Descriptive statistics and trend analysis:

The following section provides detailed descriptive statistics and a trend analysis of the academic staff complement of the University of Cape Town during the period, November 1997 to August 2013[5]. The following set of tables provides an overview of the demographic characteristics of the overall academic staff complement and of the cohorts of new hires in isolation over the 16 years under study. Due to the fact that a number of observations were dropped from the dataset as detailed in this document, the number of study participants, both full-time and part-time employees, represents a sub-sample of the total number of employees at the university. However, the number of study participants is expected to be close to the number of true employees at the university. Part of this misalignment is that some academics may work at UCT but be employed through individualized contracts and arrangements and as a result not appear on the central employee datasets. In addition, since entry and exit into the workforce is fluid throughout the year, records must either be averaged over a given year or be provided at a particular point in time. Unless otherwise noted, summary statistics are reported for December in the period 1997 to 2012 and for August in 2013. In most tables the August 2013 results are omitted, as they are not directly comparable to the other years presented.

5 The trend analysis is based on calendar time and not serial time which forms the basis of survival analysis.

A summary of employees under study is presented on an annual basis, broken down by race and gender. The table below presents an overview of total employees included in the study at year-end, and is inclusive of changes that result from new hires entering the study and from terminations and retirements leaving the study. Overall, the university has experienced a significant growth in the number of faculty it employs, especially in the number of African, Coloured and Asian or Indian employees. In addition, both the proportion and total number of females employed by the university has increased significantly. The increase in academic staff is likely a result of the large growth in student admissions that has taken place over the same period of time (University of Cape Town, 2014), (University of Cape Town, 2014), (Groundup Staff, 2015). The table reports annual average percentage changes for the three separate Vice Chancellors (VC) who have led the university between 1997 and 2013. The Ramphele period can be generalized as a time of austerity as total faculty size remained relatively constant or declined year to year. However, even under limited overall growth the number of Coloured and Asian or Indian employees experienced large annual growth. The period under VC Ndebele saw robust growth in Black and Asian and Indian employees and more limited growth for Coloured employees. The number of white employees in this period increased only marginally. Similarly, the period under the leadership of VC Price sees robust growth in both the number of Black and Asian or Indian employees but also sees moderate growth in both Coloured and White employees. While under the leadership of both VC Ndebele and VC Price the university experienced comparatively higher growth among female employees than male.

Table 1: Summary of academic employees by year[6]

Year	African	Coloured	Asian/Indian	White	Male	Female	Total
1997	39	30	15	659	546	197	743
1998	41	36	24	660	543	218	761
1999	43	40	27	628	520	218	738
2000	39	46	28	570	479	204	683
Average annual percentage change under Vice Chancellor Mamphela Ramphele							
97-00	0.2%	15.4%	25.4%	-4.6%	-4.2%	1.4%	
2001	41	51	31	593	491	225	716
2002	42	52	31	590	490	225	715
2003	47	52	34	587	492	228	720
2004	53	55	37	561	479	227	706
2005	58	59	39	571	488	239	727
2006	68	63	42	576	498	251	749
2007	68	66	46	594	497	277	774
2008	79	75	54	622	517	313	830
Average annual percentage change under Vice Chancellor Njabulo Ndebele							
01-08	10%	5.7%	8.4%	0.7%	0.8%	4.9%	
2009	81	75	50	605	499	312	811
2010	92	87	58	687	546	378	924
2011	98	92	65	705	563	379	942
2012	99	90	66	699	572	382	954
2013	110	96	73	766	622	423	1045
Average annual percentage change under Vice Chancellor Max Price							
09-13	8.1%	6.6%	10.1%	6.2%	5.7%	8.2%	

Progress towards a more diverse faculty has not been uniform across faculties. An overview of employee diversification at the faculty level is important as it provides a quantifiable measure of the effectiveness of different initiatives,

6 Unless otherwise noted all tables and figures refer only to academic employees at the University of Cape Town.

which may have been undertaken at the faculty level. Table 2 reports the share of each faculty that self-identify as white for the period 1997 to 2013[7].

Table 2: Proportion of white academic employees by Faculty by year

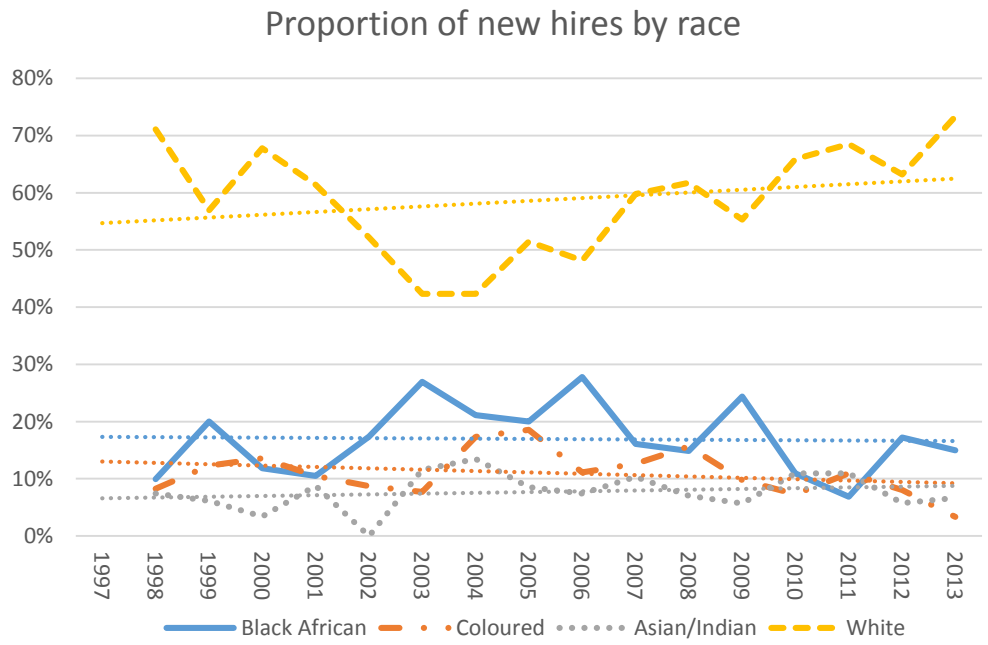
Year	Commerce	Engineering	Health Sciences	Humanities	Law	Science	Business school
1997	0.88	0.96	0.85	0.84	0.88	0.93	1.00
1998	0.89	0.92	0.84	0.81	0.86	0.92	0.87
1999	0.89	0.88	0.80	0.79	0.85	0.91	0.88
2000	0.85	0.87	0.76	0.78	0.84	0.91	0.90
2001	0.86	0.86	0.71	0.80	0.81	0.90	0.89
2002	0.84	0.86	0.68	0.81	0.81	0.90	0.90
2003	0.84	0.85	0.67	0.79	0.81	0.89	0.90
2004	0.82	0.82	0.70	0.77	0.78	0.85	0.94
2005	0.78	0.81	0.70	0.77	0.75	0.85	0.89
2006	0.78	0.75	0.68	0.76	0.77	0.84	0.79
2007	0.75	0.75	0.70	0.78	0.77	0.83	0.68
2008	0.75	0.74	0.65	0.80	0.70	0.83	0.63
2009	0.77	0.77	0.64	0.78	0.69	0.83	0.65
2010	0.77	0.74	0.67	0.75	0.72	0.83	0.63
2011	0.72	0.74	0.68	0.72	0.75	0.82	0.63
2012	0.74	0.74	0.68	0.71	0.72	0.83	0.62
2013	0.72	0.73	0.68	0.72	0.74	0.84	0.59

Diversity at the faculty level differs quite significantly. As of August 2013, 84% of academic employees in the Faculty of Science identify as white, while only 59% of academic employees in the Graduate School of Business identify as white. The other faculties lie between these two extremes. In addition, the Graduate School of Business experienced the largest percentage point change in the number of employees who identify as white, the Faculty of Science

7 Shares are based on year-end (December) cross-sectional data. Data for August is used in 2013 as I did not have access to data for September to December of 2013.

experienced the smallest. The faculties of Commerce, Engineering and the Built Environment and Health Sciences all reduced the number of employees identifying as white by at least 15 percentage points. Overall, the earlier portion of the study period (1997 – 2003) experienced more rapid reductions in the number of employees identifying as white than did the latter portion of the study period (2004 – 2013). In some cases, the percentage of white employees increased during the 2006 – 2010 period.

Figure 1: Racial composition of new hires

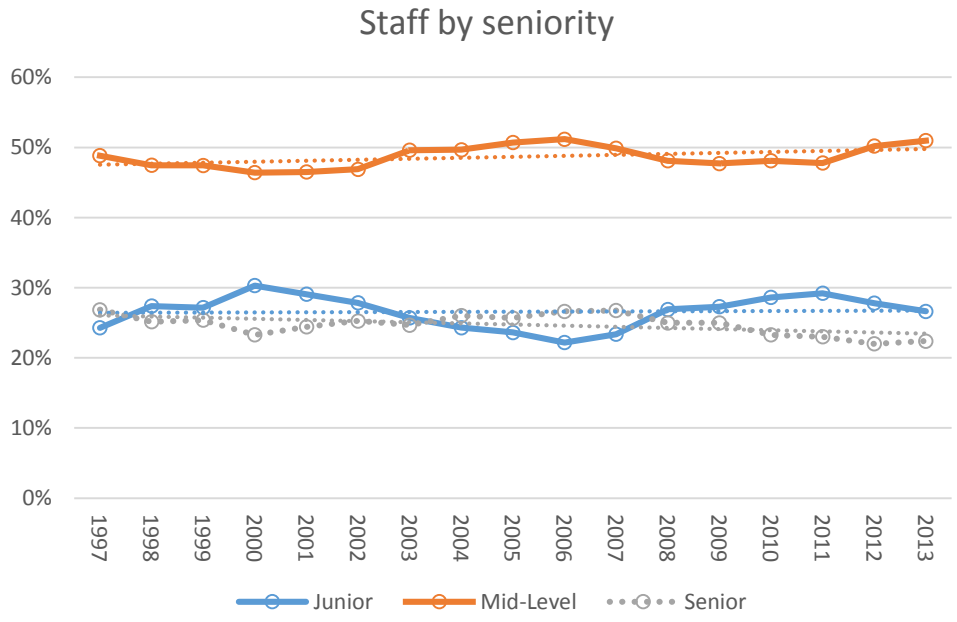


The above figure offers a breakdown of new entrants by race category. Proportions are provided as at calendar year end. White employees account for the majority of new hires at the university in most years. Between 2002 and 2006 white employees account for 50 percent or less of total new hires. In 1998, white hires accounted for in excess of 70% of new hires. This proportion decreased to below 45% in 2003, but has again risen to 70% in more recent years. The next highest proportion of new hires is accounted for by black

employees who represented approximately 10% of new hires in 1998, rising to just below 30% of new hires in 2006 and falling back down to approximately 15% of new hires in 2013. Overall, there has been a modest upward trend in the proportion of new hires that are white, in the period 2003 to 2013 after initially falling between 1997 and 2002, while other races have experienced relatively smaller changes in the number of new hires that they account for.

Figure 2 provides an overview of the rank breakdown (junior, mid-level and senior) for study participants. Mid-level employees (Senior Lecturer and Associate Professor or equivalent) account for approximately 50% of academic employees at the university, while senior employees (primarily Professors) and junior employees (typically Lecturers) account for 25% of employees each[8]. These proportions remain relatively stable throughout the study period and exhibit no meaningful trends.

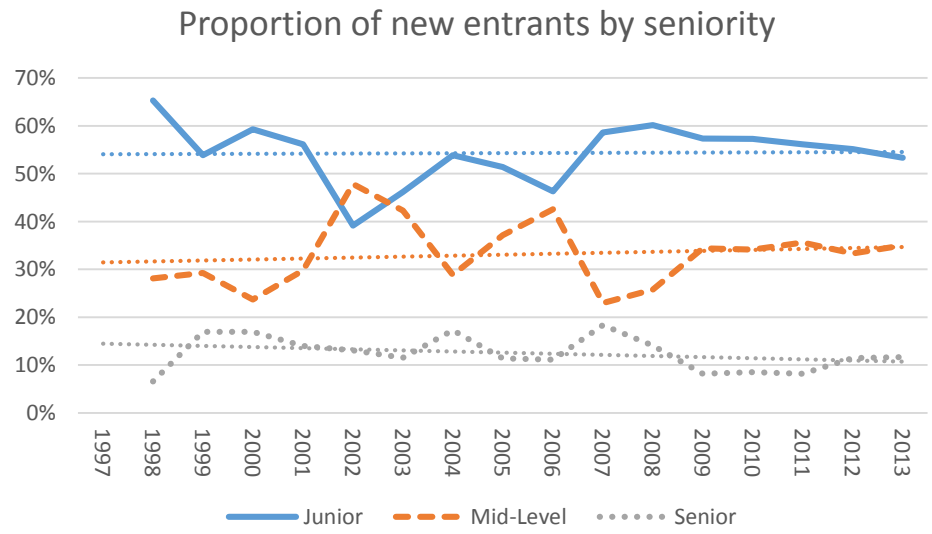
Figure 2: UCT employees by seniority



8 See table A2 in the appendix for a detailed description of the classification of employment ranks.

Among only new entrants there is high variability in the proportion of new hires per seniority category, likely representing specific hiring needs. Only 10% of employees are initially hired into senior positions and this likely reflects a practice of promoting employees already in the organization into these senior positions. The proportion of mid-level hires has fluctuated from around 25% to almost 50% of employees hired in a given calendar year. Junior hires contribute the largest proportion to total new hires (38% - 68%). Only in 2002 did the absolute number of mid-level hires exceed the number of junior new hires. The number of new mid-level hires exhibits a slight upward trend, while the number of new senior hires exhibits a gentle downward trend. The proportion of junior hires does not exhibit a meaningful trend.

Figure 3: New hires by seniority



The overall staff body has experienced significant changes in terms of racial composition over the study period. While white employees accounted for almost 90% of the academic staff body in 1997, they account for approximately 70% in 2013. This is primarily the result of modest increases in the proportions accounted for by all other races. The proportion of black employees more than

tripled from below 2% in 1997 to above 6% in 2013 while Asian/Indian employees and Coloured employees almost doubled their proportional representation. These findings are graphed in figure 4 below.

Figure 4: Overall racial composition of UCT employees

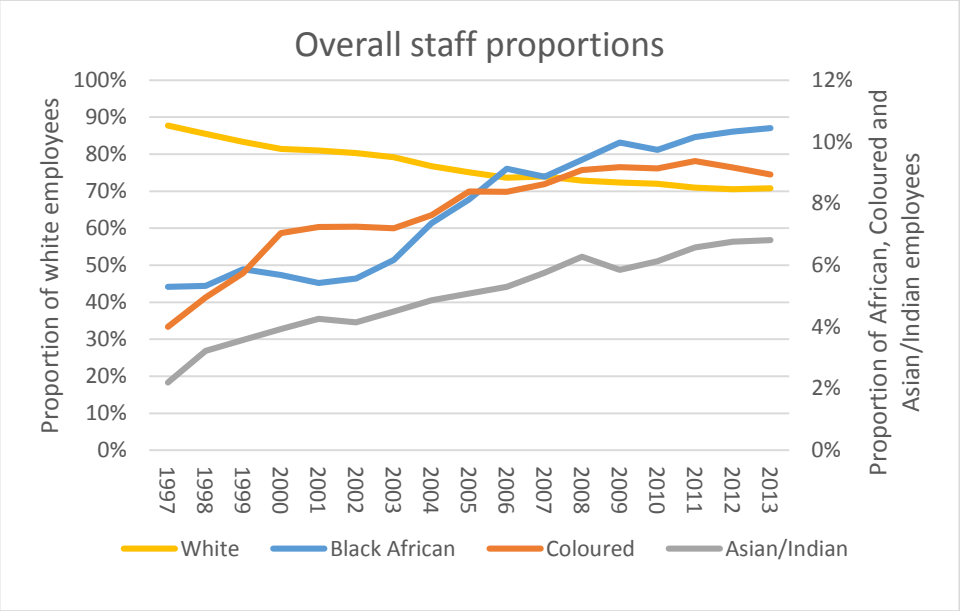


Figure 5 and 6 provide an overview of the overall gender composition of the study participants, as well as the gender proportions of new entrants on a yearly basis. UCT has traditionally been a male dominated institution. In 1997, more than 70% of its academic employees were male. This proportion has declined steadily over the study period. As of 2013, males account for fewer than 60% of all academic employees at the university. There is significantly higher gender parity when considering only new hires and in two years (2008 and 2011) the number of female new hires slightly exceeded the number of male new hires. This suggests a serious effort is being made to ensure greater gender parity in terms of academic employees.

Figure 5: Gender proportions of UCT employees

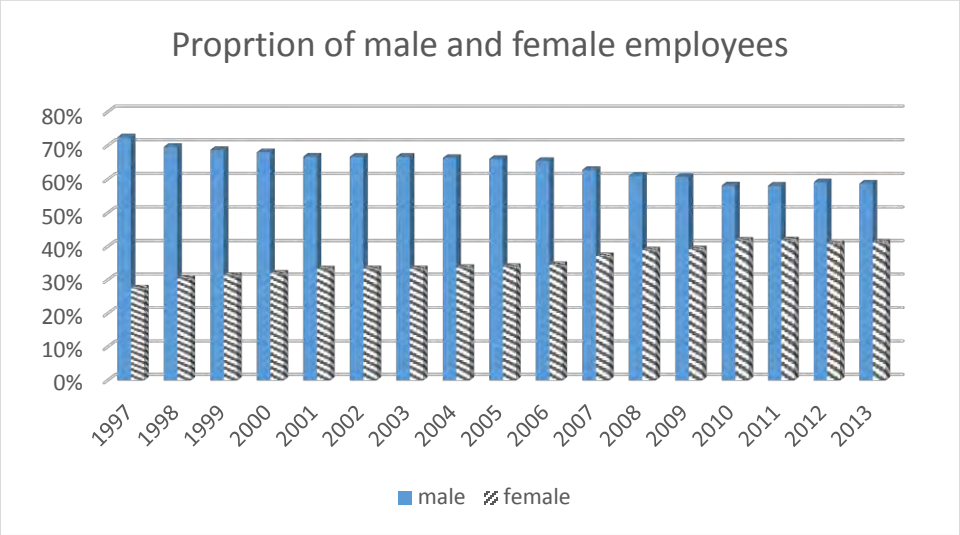
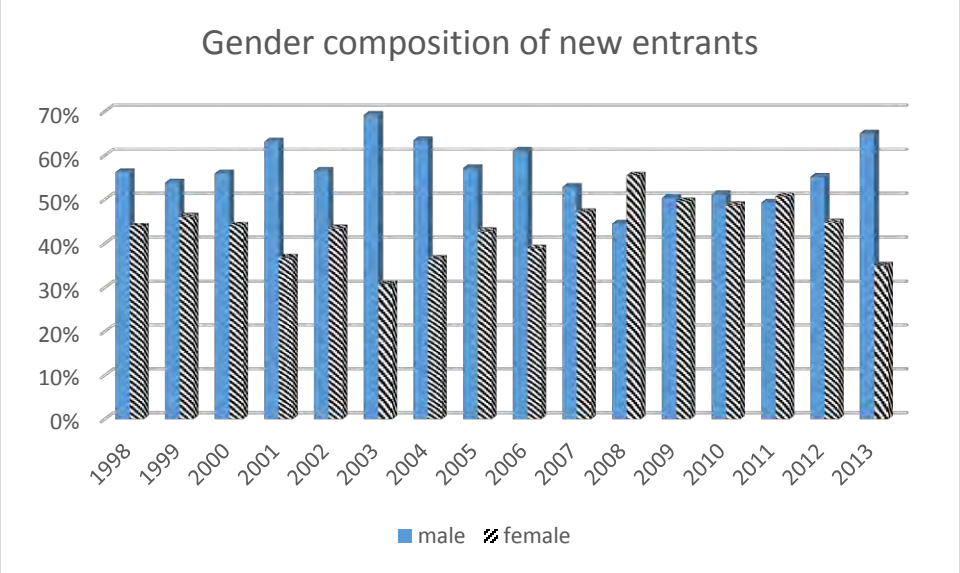


Figure 6: Gender proportions of new entrants

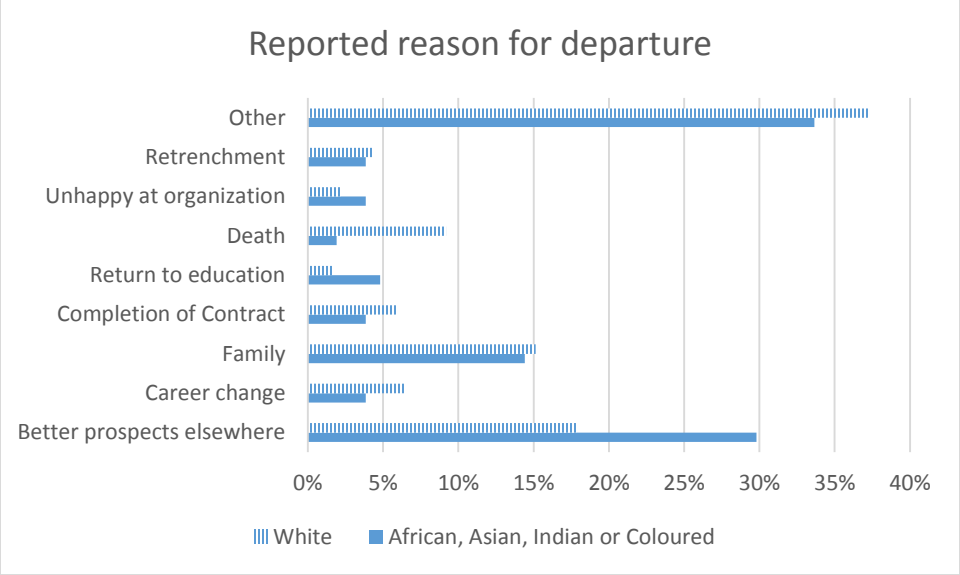


3.1 Departure

The university's Human Resources department aims to report reasons for departure from employment for non-retirement terminations. Of the 540 observed non-retirement terminations, 288 have a reported a reason for departure. While this data is incomplete and subjective, a comparison is reported since these reasons can provide insight into important non-quantifiable

explanations for differences in tenure between categorical groups. These are reported in summary separately for white employees and combined[9] for African, Asian, Indian or Coloured employees.

Figure 7: Reported reason for departure (racial split)[10]



While fewer than 2.5% of white employees cite their reason for departure as being unhappy at the organization, 4% of employees of colour provide this explanation. white employees depart more frequently due to death, but this is likely due to the combination of a higher median age for this group of employees and a small sample size. While about 2% of white employees are returning to education, almost 5% of employees of colour are planning to further their education. However, the largest difference in reported reason for departure between employees of colour and white employees is the view that better prospects exist elsewhere. Over 30% of employees of colour see better

9 African, Coloured, Asian and Indian were combined to achieve a suitable sample size. In addition, South Africa's history suggests that these races would experience the work place differently than those who identify as white.

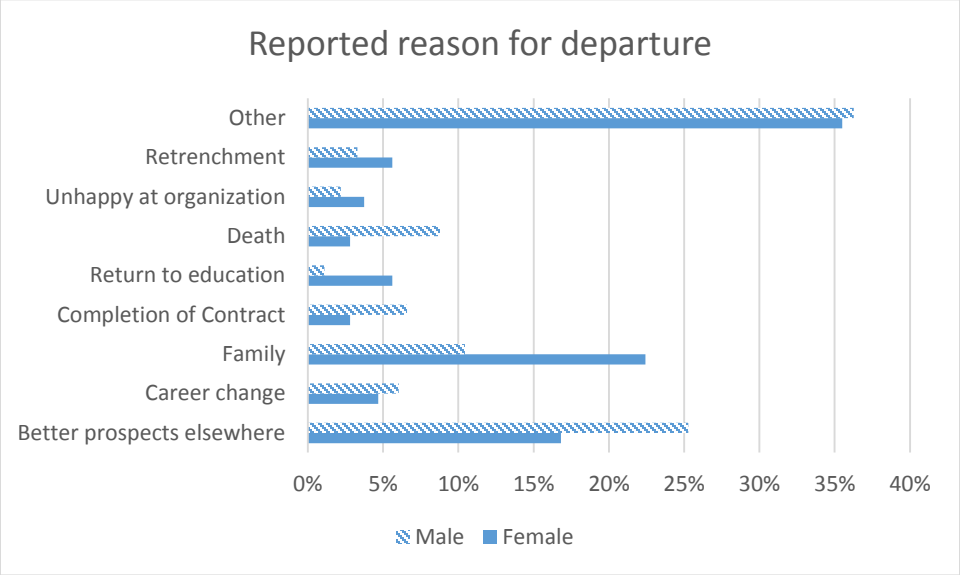
10 Categorized reason for departure based on reported reason. See Table A1 in the appendix for a detailed linking between category and reported reason.

prospects for themselves outside of employment at the university, while only 17% of white employees hold the same view. Unfortunately, I do not have access to more detailed information that could provide insight into the reason behind this difference. It seems likely that the private and government sectors offer particularly lucrative employment opportunities[11] to highly educated employees of colour. Alternatively, it may be the case that these employees face additional barriers and discrimination in promotion practices at the university when compared to their white peers.

Figure 8 reports the same data with reason for departure broken down by gender. There are a number of differences for reason reported between male employees and female employees. We see that females are more likely to be retrenched, unhappy at the organization or return to education. Expectedly, females are far more likely (11% versus 22%) to report family concerns as the reason for departure. Males much more frequently report having better prospects elsewhere as the reason for departure than do females (25% versus 17%).

[11] UCT academic salaries ranged from R528,275 for a lecturer to R967,265 for a professor all of which require at least a masters degree and slow progression from the lecturer to Professor rank (University of Cape Town, 2016). By comparison, examination of government vacancies reveals that a deputy director position in the public service requiring only a three year qualification and three years of experience has a salary of approximately R612,00 and one requiring a five years of experience may pay R726,000 (Government of South Africa, 2016).

Figure 8: Reported reason for departure (gender split)[12]



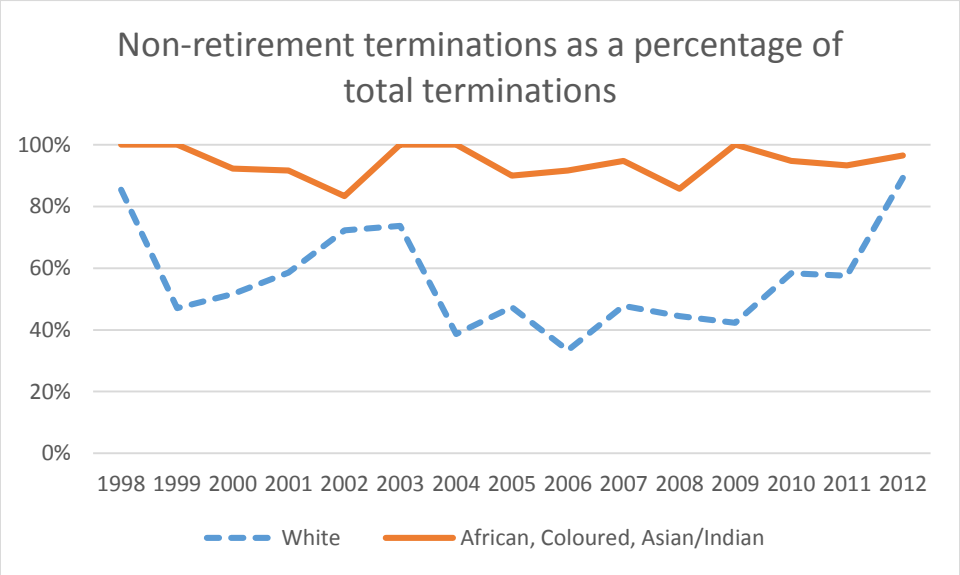
Given the high non-response rate present in the reason for leaving variable, an examination of the 252 termination that did not provide a reason for departure is useful. Non-response rates range from 45.7% for black employees to 69.1% for white employees. Females are slightly more likely to provide a response (37.2%) than are males (35%) and response rates decline with seniority. Junior employees have a non-response rate of 54.2%, mid-level employees 62.9% and senior employees 77.4%.

Figure 9 provides a visual representation of the number of non-retirement terminations as a proportion of total terminations for each race. This is informative in that there is a significant difference between white employees and employees of colour. Further, retirement terminations are of less interest, as they, for the most part, depend primarily on age and not on any particular demographic traits associated with an employee. Employees of colour exhibit a dramatically higher proportion of non-retirement terminations, particularly so in

12 See table A1 in the appendix for a more detailed description of the categorization of reason for departure.

the period from 1999 to 2009. During the 2010 – 2013 period the proportion of non-retirement terminations for white employees increases significantly.

Figure 9: Proportion of non-retirement termination



Research methodology

4.1 Motivation

I set out to decompose differences in survival between demographic groups and propose that it is not only individual demographic characteristics that determine survival, but also the semi-exogenously determined demographics of the faculty in which an employee works. That is, the interaction between members of different demographic types is hypothesized to affect focal employee survival. Under *social contact theory* we expect turnover, both voluntary and involuntary to decrease for a minority group as the proportionate size of that minority group increases. Alternatively, under *conflict theory* we would expect turnover to increase as the proportionate size of a minority group increases. For example, if the number of females in a traditionally male-dominated department increases, men may perceive their dominant position as in danger and therefore increase the level of hostility within the department.

I control for worker performance and general economic trends since these factors are likely to affect the likelihood of survival. I use a proxy for performance, namely citations per month. Further, I include a variable that measures changes in gross domestic product (GDP) and a measure of the national unemployment rate in South Africa. It is quite likely that, given the specialized nature of most university employees' work, and the relative isolation university employees experience from the rest of the economy, general unemployment and economic conditions are not strongly related to the likelihood of survival at the university. However, an expanding economy may offer a greater number of alternative employment opportunities in the private sector, particularly in highly skilled jobs. Thus UCT employees may be more incentivized to seek employment elsewhere in periods of economic expansion. Alternatively, high unemployment rates could make employees more reluctant to leave relatively secure employment at the university.

In terms of employee performance, there are two possible effects. I suspect very poor performers to face elevated hazard rates as they are unable to meet the demands of a leading academic institution and as a result will either voluntarily or involuntarily have to seek more suitable employment. However, it may also be the case that the very top performers also face increased hazard rates as they likely have good opportunities at other universities or lucrative private and government sector opportunities. Further, it is likely that termination hazard will vary with time. I suspect hazard rates to start low during the first few months of employment at the university (at which time the quality of the employee – employer match is still unknown). The hazard of failure would then increase for the first few years of employment (during which the quality of the match between employer and employee becomes apparent) and then decline after some peak hazard rate is reached. If retirements were modeled as a

hazard, I would again expect a second upturn in the hazard rate for older employees.

I suspect that the dynamic relationship between race and tenure exhibits different patterns for South African citizens and foreign citizens, given the different motivations for these groups when seeking employment in South Africa. This relationship is further complicated by a complex framework of visa regulations.

4.2 Survival analysis

In survival models, the probability of survival (continued employment at the university) at any given point is the outcome being estimated. Estimation of survival has to be done using specialized statistical techniques since traditional statistical models provide biased results in the presence of censored data. The theoretic simplicity of survival analysis makes it a particularly appealing approach to studying event history data. Three basic mathematical equations underlie survival analysis, (1) the survival equation, (2) the hazard function and (3) the cumulative hazard function[13].

$$1.\, S(t) = \Pr(T > t) = 1 - F(t)$$

$$2.\, h(t) = \lim_{\delta t \to 0} \frac{P(t \le T < t + \delta t \mid T \ge t)}{\delta t} = \frac{-d\ln S(t)}{dt}$$

$$3.\, H(t) = \int_0^t h(u)du$$

The survival function $S(t)$, represents the portion of the sample still surviving at time T. The survival function characterizes an employee's survival in terms of total length of tenure at the university. The hazard rate function, $h(t)$,

13 For an approachable overview of the theory underlying survival analysis see *Applied Longitudinal Data Analysis* (Singer & Willet, 2003)

characterizes the rate of change of the survival function and provides the likelihood of failure at any given point in time conditional on that employee still being in employment at that point in time. The hazard rate provides a measure of the risk of termination faced by an employee at specific points in that employee's tenure. The cumulative hazard function, $H(t)$, is the integral over time, of the hazard rate. The cumulative hazard function provides information about the total risk that an employee has faced throughout their tenure at the university. The primary goal of the survival analysis carried out in this study is to obtain smooth and unbiased survival estimates which is not possible with the use of conventional regression methods given the nature of the data.

The primary analysis will depend on both non-parametric and semi-parametric survival analysis methods. The non-parametric approach places no assumptions on the underlying population distribution of the data and provides simple yet powerful comparisons between groups. However, those methods do not readily allow for the modeling of covariate effects, which is more appropriately dealt with by the use of semi-parametric techniques. In the semi-parametric analysis the Cox proportional hazards model is used to model relative hazard ratios between groups while at the same time controlling for a wide range of possible explanatory variables. The nature of survival analysis data results in a sample size that decreases with time. Thus, a survival analysis is more valid (due to the larger sample size) in earlier parts of the curve, rather than in later parts of the curve (Ghaemi, 2009).

4.3 The choice between continuous time and discrete time analysis

There are two main approaches available to an investigator when modeling survival data. One is to assume that there is an underlying continuous-time model and then estimate the model's parameters by methods that take into

account the discrete nature of the data. The other approach is to assume that events can only occur at the discrete-time points measured in the data, and then apply discrete-time models and methods. In practice, both these approaches lead to very similar estimation procedures, and as such may both be described as discrete-time methods (Allison, 1982). Even though the data is presented in discrete monthly periods, the underlying process is continuous given that employees can enter into and exit from employment at any time during the calendar year. As such, the survival models used to model this should model time in a continuous manner.

4.4 A note on censoring, truncation and delayed entry

The amount of censoring in a survival analysis study depends on (1) the rate at which the event of interest occurs and (2) the length of data collection. Higher incidence rates and longer study periods result in lower rates of censoring. The UCT HR data used in this study exhibit both right censoring and delayed entry into the study. Right censoring is the most frequently encountered type of censoring in the context of survival analysis and occurs when a subject participates in a study for some period of time and is then no longer observed. In the context of this study, right censoring occurs for two reasons; (1) at the end of the study period (August 2013), all individuals who as of that date have not experienced a failure event are censored, and (2) some subjects are lost to follow up (death and retirement). In these instances I assume censoring occurs randomly and is unrelated to the reason for failure[14]. Right censoring is easily dealt with in econometric models by the use of survival analysis.

14 The assumption of non-informative censoring is required for valid survival analysis; if instead censoring is non-random and informative all analysis is invalid (Singer & Willet, 2003). There is no evidence to suggest that the censoring mechanism present in the UCT employment data is informative.

The second category of censoring present in this data set is left truncation and delayed entry, which implies a period of ignorance subsequent to the onset of risk but preceding the start of the study. This study is only able to measure risk from December 1997 onwards, however many individuals entered employment at the University at some time prior to this date. It would be incorrect to include information that relates to periods prior to the start of the study's observational period. This would introduce bias since we only have information on those individuals who survived to the start date, while we have no information on those who were employed and left employment prior to the start date of the study (left truncated observations).

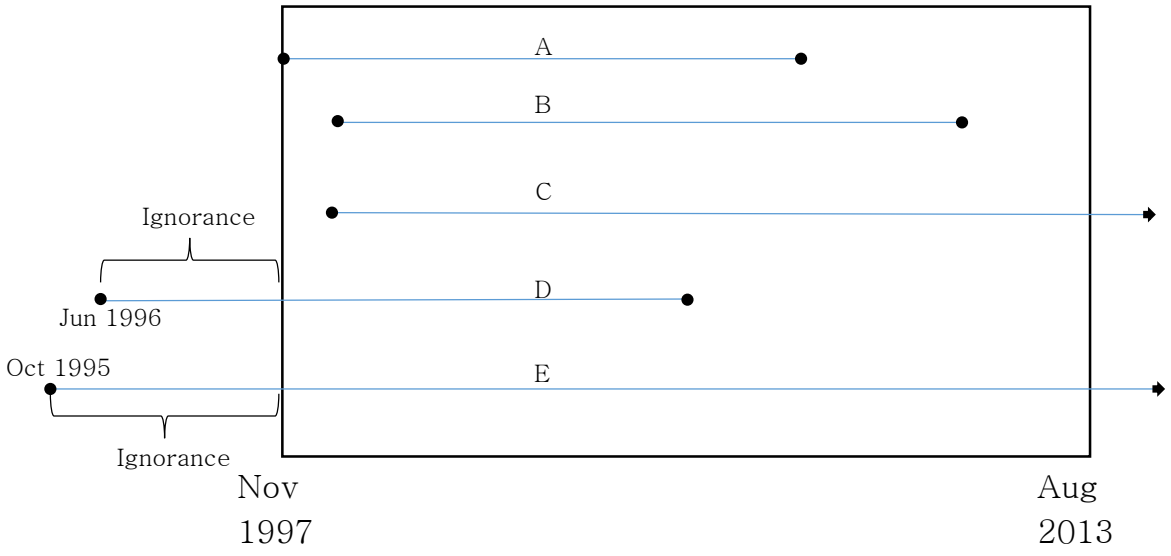
Left truncation can potentially cause serious selection bias in survival analysis because it underestimates the risk of experiencing a particular event (Liu, 2012). Left truncation occurs when an individual's survival time is not observed from some starting point in time to the occurrence of a failure event but rather, from some intermediate time tb exceeding t0 and is often observed in demographical research (Liu, 2012). Methodologically, those individuals who are left truncated should have an event time defined as T >= t | tb instead of T >= t (Liu, 2012). When left truncated observations are counted into all the risk sets from t0 to t, the hazard rate is underestimated because the denominator of the partial likelihood function is inflated by involving the left truncated observations whose left truncation time exceeds the observation time. It is relatively easy to handle left truncation if the delayed entry time tb is conditionally independent of T.

Therefore, in order to obtain the correct likelihood distribution[15] in the presence of delayed entry, one must only allow an individual to start contributing

15 The conditional likelihood contribution for a delayed entrant who enters the study at t_b and experiences a failure event at some time T > t_b is:

observations towards the analysis at time tb + 1 and discard the preceding periods, t = 1, …, t = tb (Rabe-Hesketh & Skrondal, 2012). Alternatively, bias can also be avoided by simply discarding all records related to individuals who experienced delayed entry. This procedure, however, comes at the expense of a loss in statistical power and efficiency (Rabe-Hesketh & Skrondal, 2012). I present two diagrams below in which I visualize the nature of censoring in the UCT HR data set.

Figure 10: Employees in conventional time



In Figure 10 five hypothetical employees (employee A – employee E) are used to illustrate the different censoring concerns present in my data set. The rectangular box represents the study measurement period (November 1997 to August 2013) during which I have full employee data available. Employee A and Employee B can both be handled without any adjustments as neither experiences

---

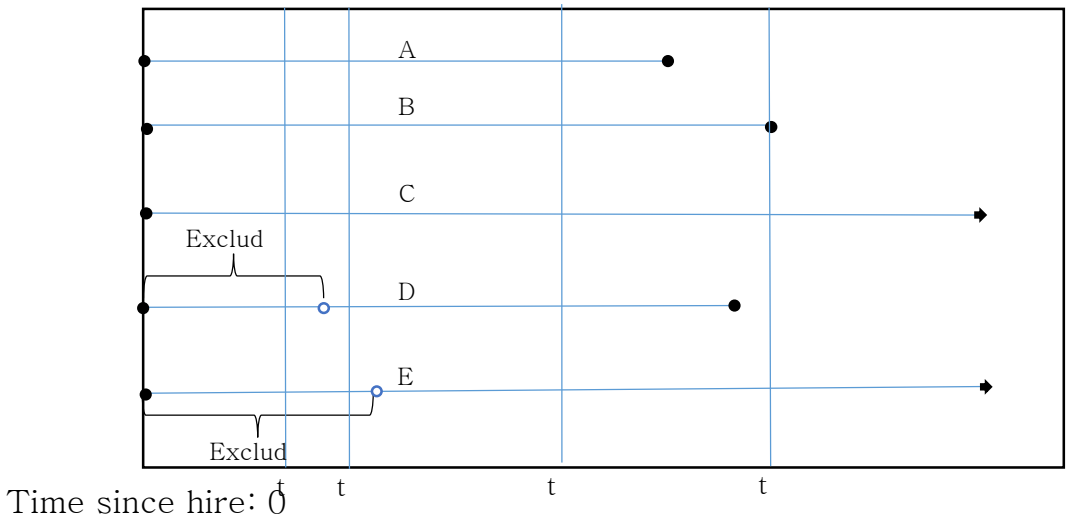
$$\Pr(T_i = t \mid T_i > t_b) = \frac{\Pr(T_i = t,\ T_i > t_b)}{\Pr(T_i > t_b)} = \frac{h_{ti}\prod_{s=1}^{t-1}(1 - h_{si})}{\prod_{s=1}^{t_b}(1 - h_{si})} = h_{ti}\prod_{s=t_b+1}^{t-1}(1 - h_{si})$$

The conclusion for the conditional likelihood function for a right censored delayed entrant follows in a similar fashion. The above equation is based on equations found on page 773 of Multilevel and Longitudinal Modeling Using Stata (Rabe-Hesketh & Skrondal, 2012).

any type of censoring. Both employees enter at or subsequent to the start of the observation period (November 1997) and both depart employment before the end of the observational window (August 2013). Employee C is said to be right censored as they enter employment during the observational window and remain in employment at the end of the observational period. We know that employee C is likely to depart at some point in the future, but not exactly when. While we know both the date of initial hire and of termination for employee D, there is a period of observational ignorance between the initial hire date and the start of the observational window. Thus, employee D is said to be a late entrant into the study – in other words that employee has already been exposed to some risk prior to the start of the study (Jun 1996 – Oct 1997). This is also sometimes referred to as left truncation[16]. Finally, employee E experiences both delayed entry into the study and right censoring at the point the observational period ends.

16 Left truncation implies a period of ignorance between the date on onset of risk and the period of the study. On the other hand, left censoring occurs if we do not know the true start date of an employee (Cleves, Gutierrez, Gould, & Marchenko, 2008).

Figure 11: Employees in serial time



In order to conduct a valid survival analysis, a number of modifications need to be made to the data set. To account for these problems I transform the dataset from calendar month to serial time (months since the start of employment). Under serial time, all employees start at time $t = 0$. A further adjustment is necessary for the inclusion of delayed entrants such that they only contribute to the study in periods subsequent to their entry into the study and not their first exposure to failure risk.

By this adjustment, I know that employees D and E entered employment prior to the monitoring window and that they experienced termination risk prior to this window. However, I do not want to include them in analysis for periods under which they were not under observation, thus I mark them as delayed entrants (the hollow circle marks their entry into the study) and their data will only be included in survival equations in periods subsequent to study entry. Therefore, in the case of employee D and employee E, their details will not contribute information to survival models to the left of the hollow circle. Employees C and

E are marked as censored on the last month under which we have them in the monitoring window.

I use four points in serial time (t1 – t4) to illustrate this. At serial time t1, employees A, B, C, D and E are all in employment, however, at this time neither employee D nor E has entered the period of observation and as such their information is not used in the construction of hazard rates at time t1. At time t2, employee D has entered the observational window, while employee E is still outside of the window. As such, hazard rates at time t2 are calculated utilizing information from employees A, B, C and D only. At time t3, observations from all five employees are within the study period and thus all five employees are used in the calculation of survival statistics. Finally, at time t4, the end of the study period, we have one observed failure (employee B), two prior failures (employees A and D), and two right-censored observations (employees C and E). The hazard ratio at this point will take into account employee B and the two censored employees.

4.5 Data overview

The following section of this thesis provides a detailed overview of the data used in this study. This is followed by a section describing in detail the preparatory data cleaning and transformation that was carried out prior to the data being used in survival analysis. A description of variables used in the economic models then ensues. All analysis and data preparation was conducted using the Stata software package[17].

Internal administrative employee records were obtained from the University of Cape Town (UCT) Human Resources (HR) department. The employee records provided covered the period of December 1997 to August 2013. The university

17 Analysis was conducted using Stata Version 13.

migrated from an earlier to a newer version of an SAP administrative database in 2004 resulting in differently formatted data for the available variables in the 1997-2003 period and the 2004-2013 period. This change reduced the number of useable variables as only those consistent in both dataset periods could be utilised. The raw administrative records are structured in a long-data format such that each line in the data set represents a unique employee event. For example, an employee's initial appointment, promotion, department change and termination would each be represented by a separate line entry in the HR administrative data set. Thus, an employee hired in 2005 that experienced no changes in their employment status is represented by a single line in the dataset. On the other hand, an employee who experienced a promotion at some point in their tenure would be represented by two lines – the first recording the initial hire and the second recording the details of the position subsequent to promotion. Birth dates and additional information on retirees were recorded in two additional administrative data sets. Information contained in these datasets were linked to the primary administrative data set using a unique employee identification number.

In addition to data obtained from UCT Human Resources department, external data sources were utilised to obtain variables for use in the econometric models. Economic indicators on the national unemployment rate and GDP growth rate[18] were obtained from the South African Reserve Bank. These variables will act as controls in the various models as it is hypothesised that economic conditions influence individual employment decisions for those working at the university. It is probably the case that an individual will be more likely to leave employment in search of external opportunities when unemployment is low and GDP growth is

18 Series KBP7019; Series KBP6006

high. Similarly, poor economic conditions might result in lower rates of turnover when alternative employment opportunities become scarcer.

Given that departmental staff performance reviews are considered confidential and were not available for use, an alternative measure of individual employee performance was developed for this study. Including a measure of individual performance in the econometric models is important since it is hypothesized that employee performance is likely to have an impact on the likelihood of an individual remaining in employment. Publicly available publication and citation count data were used for this purpose. Numerous studies suggest that both publications and citations are useful measures of academic success (Moed, Burger, Frankfort, & Raan, 1985), (Creswell, 1985), (Meho & Sonnewald, 2000). Google Scholar was used to create both a total publication count and a citation count for the period for which an individual was in employment at the university. Google Scholar was chosen over the primary alternative, Reuters Web of Science database, which offered similar functionality, for two reasons. Firstly, a free software program "Publish or Perish"[19] was available which interfaces with Google Scholar allowing for complex search criteria. This allowed me to greatly limit the number of false-positive matches and increase the efficiency of the lookup process (Harzing, 2010). And, secondly, Google Scholar is less selective in what it includes in its dataset and thus may capture a more comprehensive account on an individual's academic scholarship, while the Reuters Web of Science product is limited to a 'selection of journals' (Meho & Yang, 2007).

Using the 'Publish or Perish' software I ran an individualized search for each of the 1767 university employees to obtain those publication and citation records

19 The 'Publish or Perish' software program is available for free download on Professor Ann-Wil Harzing's website at: http://www.harzing.com/index.htm

as reported by Google Scholar. Results were limited to works published during the years an employee was employed at the University of Cape Town, and to those containing the words "University of Cape Town". This was done to further limit the number of false-positive publications. Moreover, the manual nature of the search procedure provided an extra layer of data scrutiny as I was able to compare results to an employee's academic discipline, thereby reducing the incidence of false matches. Using this data I collected both total citations and total publications for each individual. These two statistics allowed me to calculate a range of different performance measures, the primary being a measure of citations per month and a measure of the number of citations an author's papers receive on average.

4.6 Data preparation

The following section details the steps involved in preparing the data for the survival analysis. The first step consisted of data cleaning and error correction. Significant changes were required to ensure consistent spelling of names, departments and titles. The raw data contained numerous errors and inconsistencies – likely the result of poor quality control when initially recorded. Duplicate records or in some cases apparent missing entries, incomplete person entries, conflicting records and inconsistent abbreviations were all present. These errors went so far as to include inconsistencies in recording an individual's race or gender. In these instances the modal characteristics of the employee's records were used for that employee as a rule.

Following this initial preparation, the cleaned person-event data set was then converted, using a semi-automated procedure, to a monthly person-period dataset. The result is a data set containing an observation reporting the status of an employee for each month that the employee remains in employment at UCT.

While the original data set contained exact dates on which events occurred, the move to a person-period dataset necessitated using fixed periods of time. Monthly periods were chosen as a compromise between accuracy and computational feasibility, and still allowing for the use of continuous time survival models under the reasonable assumption of an underlying continuous time process. Table 3 illustrates the structure of a person-period data set for two fictional university of Cape Town employees.

Table 3: Example person-period data set

ID	Month	Year	Serial Time	Education	Race	Termination
999	Jan	2000	1	Doctorate	White	0
999	Feb	2000	2	Doctorate	White	0
999	Mar	2000	3	Doctorate	White	1
101	Dec	2003	1	Master's	Black	0
101	Jan	2004	2	Doctorate	Black	0

During the conversion procedure, further errors in the dataset became apparent. In numerous cases, the end date of an individual's record for a particular status overlapped with the start date of another event for that same individual. For example, an employee may have been recorded as holding the position of "Senior Lecturer" in years 2001 to 2005 while a separate record shows that same employee as holding the position of "Associate Professor" in years 2003 to 2008. As a rule, it is assumed that the start date is more accurate than the end date of an event – this is based on the fact that newer records likely replaced older records that had not been properly updated. In cases of an apparent missing record, the most recent known information was used to fill in the missing information. This was accomplished using a sequential, 'carry forward', and 'carry backward' routine in Stata. An individual's status was 'carried forward' until a subsequent status update was encountered. This was

followed by carrying an employee's status backwards until an earlier record is encountered, or the start date of the study was reached. For example, if an individual has education recorded as "Master's degree" in October 2003, and a new record updates education to a "Doctoral degree" in September 2006, the carry forward routine will carry forward the "Master's degree" to all periods from October 2003 up until and including August 2006. Where the final year worked did not agree with the date of retirement recorded in the retirement data set, the retirement data set was assumed to be correct. In these cases, the final month of full employment was set equal to the start date of retirement.

After harmonization, the combined dataset contained individual level variables covering the following information for academic employees: age, national origin, race, academic qualification, gender, department, faculty, initial hire date, promotions and positional changes, salary pay grade, title, and reason for withdrawal[20] if such occurred. The final monthly person-period dataset contains 150,152 records providing information on 1767 unique employees. Of these 683 were actively employed by the University at the start of the study, while the other 1084 employees entered into employment at the University at some point during the period under study.

4.7 Variable definitions

The primary dependent variable under study is length of tenure. This variable is measured as the number of months from the date of initial hire at the University of Cape Town. The definitions of other key variables used in the survival analysis are provided below.

20 The reason for withdrawal is not always provided and a fairly wide range of responses were allowed necessitating the grouping of these responses into distinct categories. See Appendix 1 for a detailed overview of the classification of employee responses.

Race is a categorical variable with five categories, (1) white, (2) African, (3) Coloured, (4) Asian or Indian and (5) unknown[21]. These categories were chosen as they follow the typical racial classifications in use in South Africa and UCT employees were able to report their self-identified race using these categories. Asian and Indian were reported as two separate categories in the raw data but the decision was made to combine the two categories as each constituted only a very small sample and there is evidence to suggest that those employees falling within these two categories face similar probabilities of survival. Both South African nationals and foreign citizens are invited to report their race. To account for the different position South African nationals and foreign citizens find themselves in the South African labour market, I construct a separate dummy variable to indicate whether an individual employee is a South African citizen or not.

An indicator for the faculty of employment was created. Individuals who were reported to have less than a bachelor's degree were dropped from the study as they represent a very small number of individuals and atypical for academic university employees[22]. Faculty is used instead of the much more detailed department of employment for three reasons: (1) sample size at certain departments is very small, (2) the poor quality of data recording in the UCT HR datasets meant that reconciling departmental classifications would not be possible given time constraints and possibly only result in marginal gains and (3) it is likely that the department is too small a unit of analysis as significant cross department collaboration takes place within a university setting. However, even when using faculty and faculty equivalent grouping, those employed in Libraries

21 Observations reporting unknown race are dropped from the sample as this is a primary variable under study. This results in a loss of 4376 monthly observations representing 54 employees.

22 See Appendix item 4 for a detailed table of the education variable classification.

and University Administration remain problematic, as they are not typical academic departments and present a very small sample size. These employees were excluded from the analysis[23]. The largest number of observations are for the faculty of Humanities and the smallest number for the Graduate School of Business.

In addition to raw citation counts I create three categorical variables related to citation and publication counts. Indicator variables for poor performance, high performance and very high performance are created. These are defined in a comparative manner at the level of the faculty for each day in the calendar year. For each time and each faculty the distribution of citation counts is determined. The poor performance variable is defined such that it takes the value '1' if an employee achieved a count of zero citations per month. High performance is defined such that the variable assumes the value '1' if an employee's citation-per-month count falls between the 50^{th} and 80^{th} percentiles of their faculty's citation-per-month distribution for that period of time. Finally, very high performance is defined such that it takes the value '1' if an employee's citation-per-month count is above the 80^{th} percentile of the faculty distribution at that time. Separation by faculty is done to account for the fact that certain academic disciplines are much more publication heavy than others. Academics in the Faculty of Science and the Faculty of Health Sciences achieve significantly higher publication and citation counts than academics in the Faculty of Humanities or Faculty of Commerce.

Dissimilarity measures are constructed from the point of view of each individual employee and closely follow the methodology proposed by Kurtulus (Kurtulus, 2011). Dissimilarity measures are created to account for differences in

23 See Appendix item 3 for the final list of faculties included in the study.

educational attainment, positional seniority, gender, racial diversity and age differences among employees in a given faculty. These measures are calculated for each employee and, for every point in time and, as a result, vary on a monthly basis. They are intended to capture possible effects that result from changes in the faculty composition. The first set of employee dissimilarity measures are constructed for categorical variables. *Employee gender dissimilarity* is defined as the share of employees in the focal employee's faculty of employment who are of a different gender to that employee. For example, if the focal employee is female and 80% of all employees in that faculty are male, then the reference employee's gender dissimilarity measure is 0.8[24]. *Employee race dissimilarity* is constructed in a similar fashion and provides a measure of the share of employees in a given faculty who belong to a race other than the focal employees own race (White, African, Asian/Indian, Coloured). The *employee education dissimilarity* indicates the share of employees in the focal employee's faculty who have a different level of educational attainment to the reference employee (bachelor's degree, master's degree, or a doctorate). The *employee age dissimilarity* measure and the *employee tenure dissimilarity* measure are constructed differently as they are constructed from continuous variables. *Employee age dissimilarity* is defined as the absolute difference between the natural logarithm of the age of the reference employee and the natural logarithm of the average age of all employees in a given faculty. The summary statistics for key variables utilised in this study are presented below in table 4. Importantly, these present the mean value over all person-month

24 For example, a black employee works in a faculty that contains 8 Black employees, 2 Coloured employees and 20 white employees. In this case the total faculty size is 30, of which 7 are the same race as the focal employee. This results in a racial dissimilarity measure of $^{22}/_{29} = 0.759$

observations for each variable instead of the mean value of each variable for each person.

Table 4: Summary statistics for key variables[25]

Variable	Mean	Min	Max
Age	46.96	21	70
Faculty of Commerce	0.129	0	1
Faculty of Engineering	0.134	0	1
Faculty of Health Sciences	0.185	0	1
Faculty of Humanities	0.247	0	1
Faculty of Law	0.059	0	1
Faculty of Science	0.220	0	1
Graduate School of Business	0.026	0	1
Bachelor's degree	0.096	0	1
Master's degree	0.315	0	1
Doctoral degree	0.586	0	1
Junior	0.265	0	1
Mid-level	0.493	0	1
Senior	0.242	0	1
South African national	0.798	0	1
Male	0.640	0	1
Black	0.085	0	1
Coloured	0.079	0	1
Asian/Indian	0.056	0	1
White	0.780	0	1
Delayed Entrant	0.518	0	1
Unemployment	24.56	21	29.3
Δ Gross domestic product	3.03	-6.1	7.4
Citations per month	2.287	0	1941
Poor performance	0.335	0	1
High performance	0.301	0	1
Very high performance	0.197	0	1
Promotion 12 months	0.036	0	1
Promotion 12 - 24 months	0.031	0	1
Promotion 24 - 36 months	0.025	0	1

[25] Sample size N is 150152 for all variables

Dissimilarity measures were constructed such that their value ranges between 0 and 1 inclusive. A dissimilarity score of zero would indicate all employees within the faculty are of the same type as the focal employee. Alternatively, a dissimilarity score of 1 indicates that all employees in the faculty differ in type to the focal employee. The dissimilarity score depends upon both the characteristics of the focal employee and the modal characteristics of the faculty. The average dissimilarity scores experienced by academic employees at the University of Cape Town are reported in the table below.

Table 5: Summary of dissimilarity measures

Measures of focal employee dissimilarity[26]				
Variable	Mean	Median	Minimum	Maximum
Educational dissimilarity	0.509	0.500	0	0.976
Dissimilarity in seniority	0.619	0.649	0	0.872
Gender dissimilarity	0.435	0.400	0	0.885
Racial dissimilarity	0.393	0.245	0	0.976
Age dissimilarity	0.208	0.187	0.003	0.719
Tenure dissimilarity	1.402	1.033	0	5.268

The mean values of focal employee dissimilarity are of interest in themselves as they provide a measure of how much the average focal employee differs from those he or she works with. The average focal employee works with someone who is likely of the same race as them, but on the other hand they are likely to be working with others who differ in terms of educational attainment and gender. Since tenure and dissimilarity and age dissimilarity were calculated from continuous rather that discrete data they are not directly comparable to the

26 Dissimilarity measures are reported as a mean of the mean individual value. In other words, the dataset was first averaged by employee and then the mean of the individual employee averages was taken. This was done to avoid averages being biased by individuals who survive in the dataset for a longer period of time.

other measures of focal worker dissimilarity. Employees on average differ more from one another in terms of their tenure length at the university than their age.

In Appendix 7 I present a life table that tracks event histories for all employees. Failure occurs when an individual's employment is terminated for a reason other than death or retirement. The table presents the monthly person-period data grouped into 24 month periods. The lost to study column provides a count of individuals lost to censoring in each study year. The survival column reports a point estimate of the proportion surviving after a given number of months. The life table provides a representation of the study risk set at specific points in time.

4.8 Non-parametric analysis:

Under the Kaplan-Meier Product Limit estimator, each subject is fully characterized by three variables: (1) their serial time, (2) their status at the end of their serial time, and (3) the study group that they are in (Kaplan & Meier, 1958). Serial time is arranged from shortest to longest, regardless of the true calendar date at which a particular individual entered the study. By this adjustment, all members within a given group begin the study at the same point and all are surviving until something happens to one of them. An interval in Kaplan-Meier analysis is the time between entry and the point of termination and is graphed as a horizontal line – only event occurrences define known survival time intervals. The shape of the Kaplan-Meier survival curve gives an indication of whether hazard rates are proportional for different groups – in this case the Kaplan-Meier curves for different groups would exhibit roughly the same shape (UCLA, 2014). Confirming the proportional hazards assumption provides support for the use of the Cox Proportional Hazards model.

The Kaplan-Meier curves represent the portion of the initial sample surviving after a given number of months. Death and retirement were not considered terminations, as these are independent of individual performance or faculty employment decisions. Instead these individuals will be considered censored at the time of their departure. No distinction was made between voluntary and involuntary termination. This is in part due to incomplete data on the reason for termination and in part due to the issue that identification is complicated by the fact that individuals may choose to undergo voluntary termination if they expect to be fired or retrenched in the near future. Since this analysis is based on serial time for each individual, time and cohort effects are not taken into account in Kaplan-Meier analysis.

An alternative procedure of non-parametric analysis utilizes the Nelson-Aalen estimator which is used to estimate the cumulative hazard rate function from censored data. While the cumulative hazard rate function is of some interest, the hazard rate itself is the primary concern. The slope of the cumulative hazard function provides insight into the underlying hazard rates. The Nelson-Aalen estimator has a strong justification in terms of the theory of counting processes and is preferred to deriving the cumulative hazard function from the Kaplan-Meier survival function. The cumulative hazard rate curves estimated are included for reference purposes in Appendix 5.

4.9 Cox proportional hazards model

The Cox Proportional Hazards model has widely established itself as the preferred method for modelling time to event data, particularly in the presence of covariates. While Kaplan-Meier curves provide informative visualizations of the survival curves experienced by categorically defined groups, the Cox Proportional Hazards model allows for the inclusion of multiple covariates – both

continuous and discrete. The main purpose of the traditional Cox model is to estimate hazard rates in an unbiased manner (Royston & Lambert, 2011).

Under the Cox model the hazard ratio is estimated by considering each time at which an event occurs. When estimating the hazard ratio over the complete follow-up period, the same weights are given to very early hazard rates, which affect almost everyone in the sample, and to very late hazard rates that affect only the few remaining survivors. The hazard ratio is thus averaged over event times. In the case of proportional hazards, the overall hazard ratio is not affected by this weighting procedure. If on the other hand the hazard ratio changes over time – that is the hazard rates are not proportional, then equal weighting may produce biased results. For this reason is it important to test for the validity of the proportional hazards assumption.

Results

I begin with a non-parametric assessment of survival utilizing primarily Kaplan-Meier estimation. I use the results from this investigation to inform the multivariate models I specify under the Cox regression.

5.1 Kaplan-Meier analysis

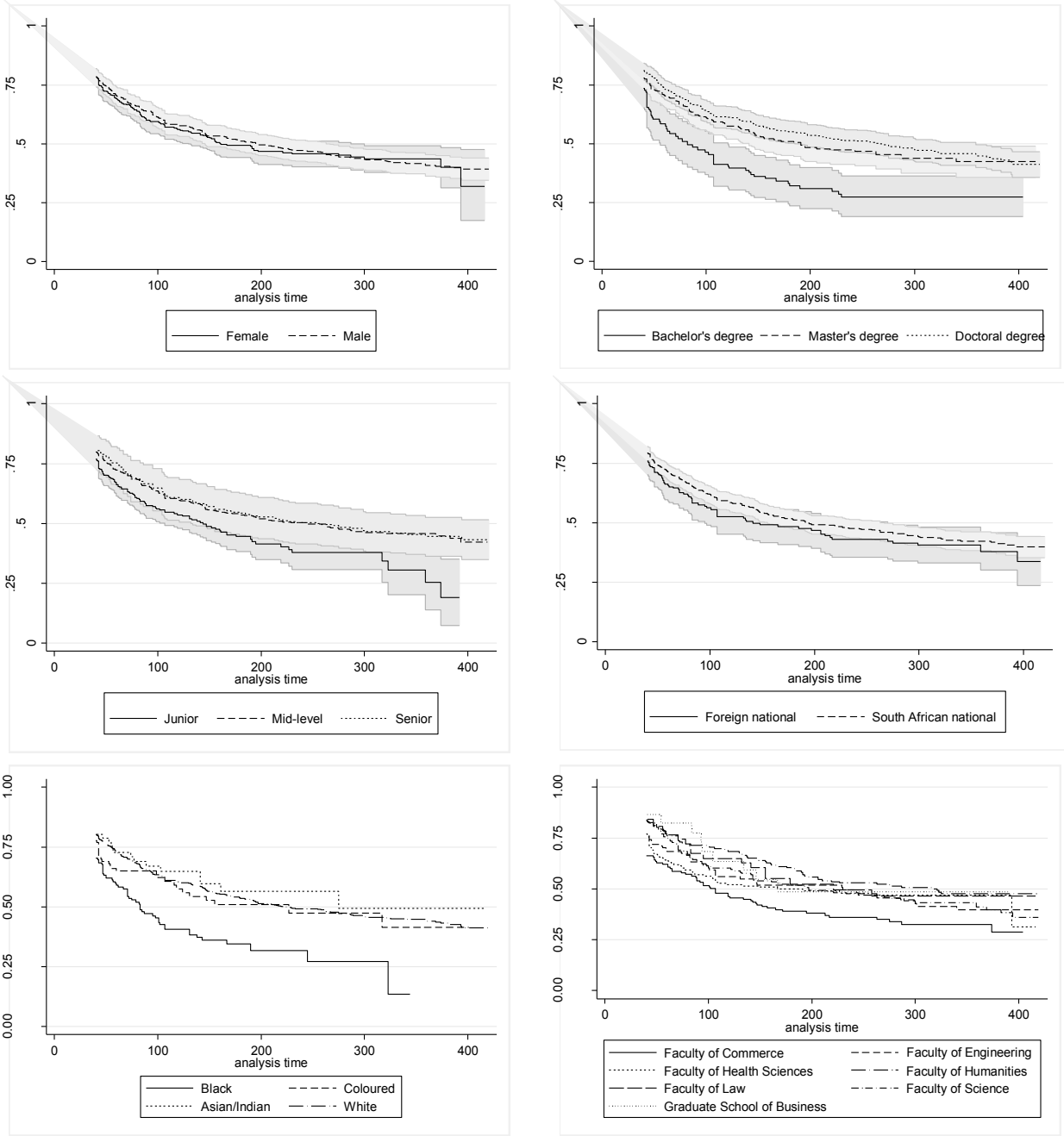
In figure 12, I presents the Kaplan-Meier product limit survival curves for various demographic sub-samples[29]. For robustness, the same exercise was conducted using a restricted sample containing only individuals who did not experience left truncation – that is their true start date corresponds to the period subsequent to the start of the study in November 1997 – this are available in Appendix 6: Restricted sample Kaplan-Meier survival curve estimates. In all cases, this restricted sample produces results consistent with

29 Confidence bands provide a visual reference of the 95% confidence-level interval. Kaplan-Meier product limit estimates are based on the full sample of all University of Cape Town employees included in the study.

the primary sample. However, as a result of the reduced sample size, confidence intervals are wider.

Educational attainment (holding more than a Bachelor's degree) and seniority (Mid-level or Senior) are both strongly associated with an increased probability of survival. There are large measurable differences in turnover between races. Black employees experience significantly higher rates of turnover than any other race. Predictably, South African nationals exhibit lower levels of turnover than foreign nations, likely the result of the different employment dynamics faced by these groups. However, this result is not statistically significant. When considering faculty level differences, the Faculty of Commerce fares significantly worse in terms of survival probability than other faculties.

Figure 12: Kaplan-Meier survival curve estimates[30]



30 95% confidence-level bands are presented in the first four graphs. The bands were omitted from the final two graph for visual purposes owing to the number of categories displayed in those graphs.

While smoothed hazard rate curves show significant upturn towards the later periods of serial time, not much emphasis should be placed on these results given the very small sample sizes involved at long tenure lengths.

Having only a bachelor's degree is associated with significantly higher hazard rates in early periods of employment as is employment in a junior position. Black associated with elevated hazard rate. The results from the hazard rate analysis are consistent with the Kaplan-Meier survival curves.

While visual inspection of Kaplan-Meier function curves can provide insight, it is important to test for statistically significant differences between the survival curves of different groups. The Peto-Peto-Prentice test[31] for equality in survival function is chosen. The results from these tests are reported in Table 6 below.

31 The Peto-Peto-Prentice test was used as it remains valid even if the proportional hazards assumption is violated.

Table 6: Testing for difference in survival curves

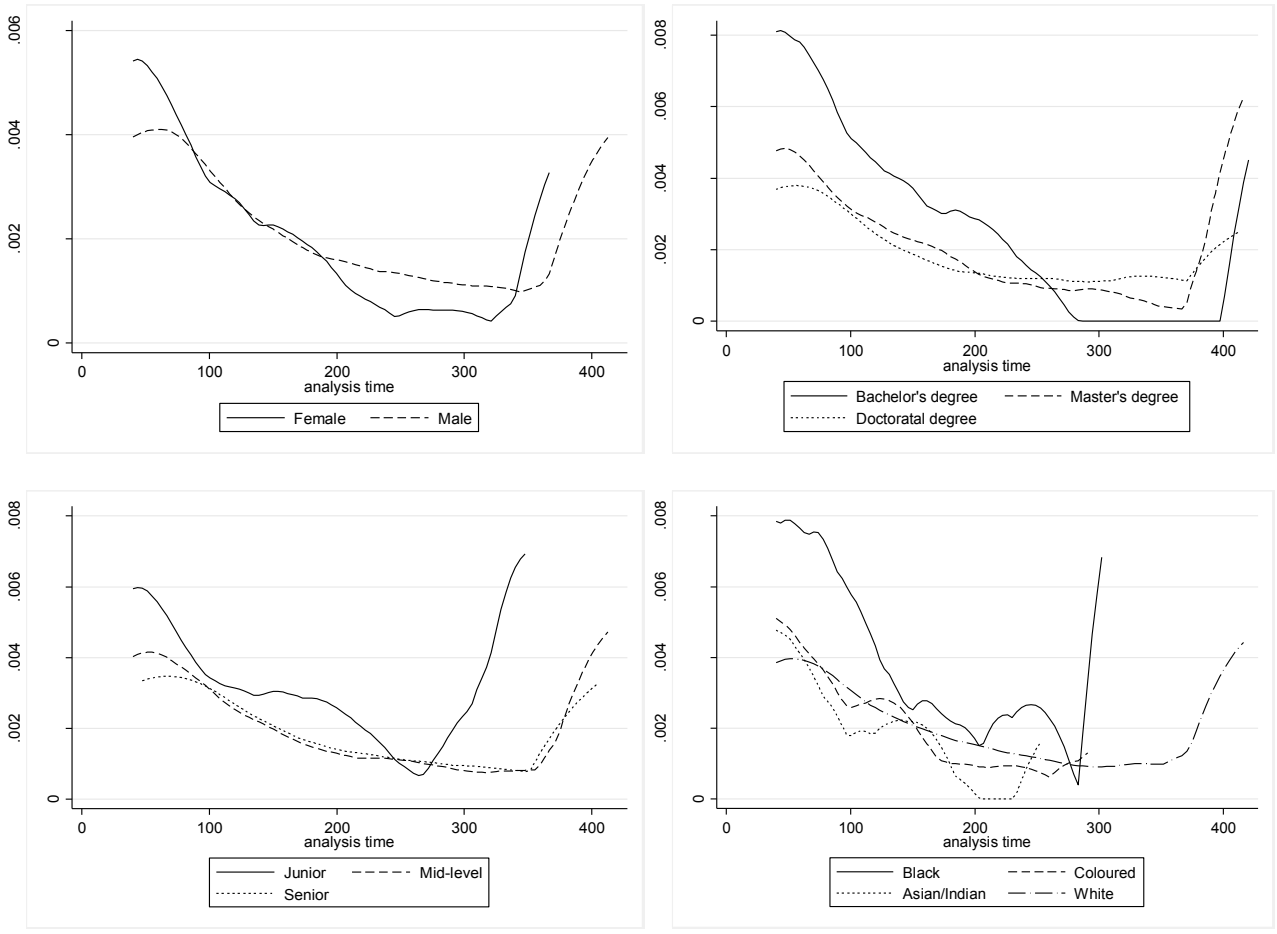
Peto-Peto test for equality of survivor functions					
	Events observed	Events Expected	Sum of Ranks	Chi2	p-value
Black	96	61.09	27.024	23.09	0
Coloured	52	51.66	0.375		
Asian/Indian	29	35.72	-5.22		
White	331	259.54	-22.18		
Female	214	210.27	3.896	0.22	0.6375
Male	294	297.73	-3.896		
Bachelors	86	55.94	22.354	18.51	0.0001
Masters	177	177.91	-0.508		
Doctorate	239	268.15	21.846		
Junior	218	187.53	21.565	7.35	0.0253
Mid-level	210	230.02	14.446		
Senior	79	89.45	-7.12		
Foreign national	123	109.71	11.437	2.66	0.1031
SA national	385	398.29	11.439		
Faculty of Commerce	100	73.15	20.898	19.55	0.003
Faculty of Engineering	73	70.43	1.963		
Faculty of Health Sciences	116	110.02	7.875		
Faculty of Humanities	88	109.17	17.128		
Faculty of Law	26	31.33	-4.167		
Faculty of Science	93	98.94	-6.611		
Graduate School of Business	12	14.96	-2.83		

Survival curves in terms of racial categorization, educational attainment and faculty of employment are all significantly different at the 1% level. Survival in terms of seniority of position is significant at the 5% level, while survival rates do not differ significantly between male and female employees, or South African and foreign nationals.

To obtain an understanding of when in time an employee is most at risk of termination I plot smoothed[32] hazard functions. This information is useful as it could be used by policy makers to target interventions aimed at increasing retention at specific periods on an employee's tenure. In general, employees experience similar shaped hazard rates which reach an initial peak fairly early in an employee's tenure and then begin to decline quite rapidly. These findings confirm the result supported by prior literature that those with lower levels of educational attainment experience higher rates of turnover. However, in terms of gender differences females employed at the university do not appear to face significantly greater hazard rates. Among racial classifications, Black employees experience significantly higher hazard rates, especially in the initial years of employment than any other race.

32 Kernel smoothing was applied to the step-wise functions in order to obtain hazard rate estimates.

Figure 13: Smoothed hazard rate estimates



I provide a number of estimates of the cumulative hazard rate using the Nelson-Aalen estimator in Appendix 5. The slope of the cumulative hazard function provides information about the hazard rates. Visual inspection of these graphs supports the hypothesis that it is during the initial years of employment at the university that an employee is most at risk of departure. After the first few years, hazard rates decrease and stabilize at a low level. At very late stages of tenure, hazard rates once again begin to increase. This supports the results obtained through Kaplan-Meier estimation.

5.2 Semi-parametric Cox Proportional Hazards estimation

All Cox proportional hazard models are estimated via maximum likelihood. Importantly, coefficients reported are not the standard regression coefficients but rather hazard rates associated with a 1 unit change in the variable while all other covariates are held constant[33]. This representation of the coefficients results in straightforward interpretation. A hazard rate coefficient of 1.2 on a variable implies that a one unit increase in that the variable will result in a 20% increase in the hazard rate. Alternatively a hazard rate coefficient of 0.6 implies that a 1 unit increase in that variable will result in a 40% decrease in the hazard rate. Furthermore, Cox models estimate the effect of covariates on the hazard rate while leaving the baseline hazard rate unspecified. As such, the models focus on relative risk rather than absolute risk. Interpretation of hazard rates for continuous variables follows in a similar fashion. For example if unemployment had a coefficient of 1.04, then an increase of one percentage point in the unemployment rate would result in a 4% increase in the hazard ratio. In this case, a two percentage point increase in the unemployment rate would result in a $(1.04)^2 = 1.0816$ increase in the hazard rate.

Table 7 reports the proportional hazard estimates, transformed standard errors[34] and associated significance level for three different model specifications. Sample A consists of the full at-risk sample of 1767. The restricted sample B is included as a robustness check and only includes only the 1094 employees who entered into employment subsequent to November 1997. This is done to ensure that inclusion of the delayed entrants is not causing unexpected bias in the results. Overall, results across the two samples are very similar for all variables.

---

33 For example, a raw Cox regression coefficient of 0.09 is associated with a hazard rate of 1.094 through the following relationship: $HR = e^{0.09} = 1.094$

34 Standard errors are converted by Stata using the Delta method (Cleves, Gutierrez, Gould, & Marchenko, 2008).

Where differences do exist they are primarily limited to a marginal loss of precision in the reduced sample. In addition there are two noticeable differences, (1) the effect on the hazard rate for black employees declines from approximately 75% to 54% while the impact of a doctoral degree increases from -35% to -45%. Sample 1 contains 502 failure events while sample 2 only sees 323 failure events. These findings suggest that the full sample, which allows for greater statistical precision, is the preferred sample and will be used though this investigation.

Table 7: Cox Proportional Hazards Models 1

	Full sample			Entry at t_0		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Black	1.750***	1.752***	1.792***	1.537***	1.543***	1.572***
	(0.237)	(0.234)	(0.248)	(0.251)	(0.252)	(0.267)
Coloured	1.125	1.125	1.113	1.149	1.150	1.114
	(0.196)	(0.175)	(0.175)	(0.204)	(0.203)	(0.198)
Asian/Indian	0.945	0.942	0.943	0.962	0.954	0.907
	(0.167)	(0.187)	(0.183)	(0.206)	(0.205)	(0.213)
Male	0.948	0.949	0.979	0.894	0.896	0.907
	(0.140)	(0.097)	(0.102)	(0.112)	(0.113)	(0.117)
South African	0.878	0.880	0.853	0.861	0.867	0.835
	(0.120)	(0.105)	(0.103)	(0.133)	(0.134)	(0.131)
Mid-level	0.842	0.843	0.837	0.914	0.919	0.996
	(0.097)	(0.100)	(0.105)	(0.132)	(0.132)	(0.155)
Senior	0.930	0.931	0.941	0.972	0.971	1.072
	(0.172)	(0.170)	(0.174)	(0.244)	(0.244)	(0.270)
Masters	0.732**	0.732**	0.719**	0.734*	0.737*	0.717**
	(0.102)	(0.098)	(0.097)	(0.117)	(0.117)	(0.114)
PhD	0.647***	0.645***	0.617***	0.554***	0.554***	0.520***
	(0.107)	(0.095)	(0.090)	(0.101)	(0.101)	(0.093)
GDP		1.065*	1.062*		1.080*	1.074*
		(0.039)	(0.039)		(0.044)	(0.043)
Unemployment		0.851***	0.857**		0.768***	0.773***
		(0.054)	(0.054)		(0.065)	(0.065)
Poor performer			2.378***			2.300***
			(0.406)			(0.495)
High performer			1.441**			1.540*
			(0.259)			(0.345)
Top performer			2.849***			2.415***
			(0.510)			(0.570)
Citations / month			1.008***			1.013**
			(0.001)			(0.007)
(Citations / month)^2			0.999***			0.999
			(0.000)			(0.0000)
Number of Failures	502	502	502	323	323	323
Employees	1767	1767	1767	1085	1085	1085
N	149722	149722	149722	71965	71965	71965

Notes: Tied employment terminations are handled using the Efron approximation[1]. All models include a full set of dummy variables for year of termination, faculty of employment and age in

years. Robust standard errors, clustered by employee id, are reported in parenthesis. Statistical significance indicated by *** (0.01 level), ** (0.05 level) and * (0.10 level).

The most basic specification (Model 1) considers the influence of simple demographic characteristics on the hazard rate experienced by an employee. Model 2 expands upon Model 1 by including two control variables intended to control for external economic factors (GDP and Unemployment). Model 3 includes variables that control for employee performance. To measure performance I use three dichotomous variables (poor performance, high performance and very high performance) and two continuous variables – citations per month and the square of citations per month since I hypothesized that both poor performing employees and high performing employees face increased risk. The two continuous measures are based on unadjusted individual citations whereas the three performance dummy variables measure an employee's performance relative to their faculty peers.

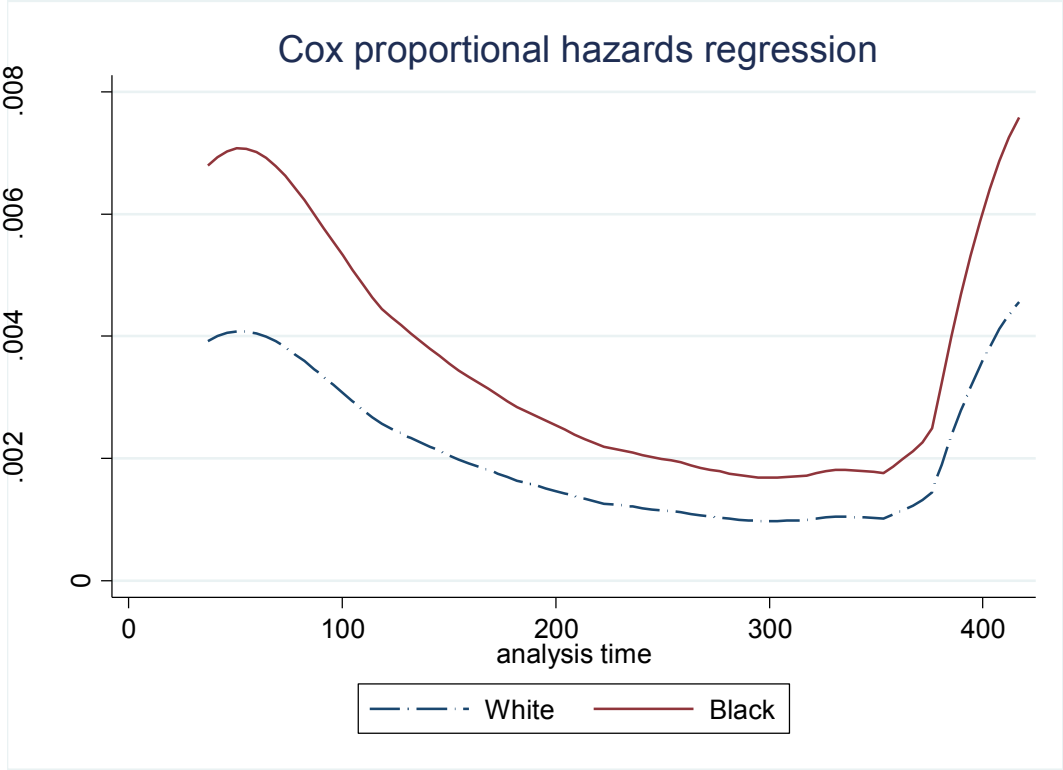
Model 1 is significant at the 1% level for Black and Doctoral degree and at the 5% level for a Master's degree. Seniority, male gender and South African nationality reduce proportional hazard but these effects are not statistically significant. The interpretation of the hazard on Black is that holding all other characteristics included in the model constant, a black employee faces a hazard rate approximately 75% higher at any given point in time that a white employee (omitted category).

In model 2, I include change in gross domestic product growth, and the monthly unemployment rate. Both variables influence the hazard rate in the expected direction. Higher rates of growth in GDP and lower unemployment rates result in a higher hazard ratio. Model 3 is the first model specification to include controls for individual employee performance. This is done primarily to test whether

racial differences in hazard rates remain even after controlling for individual performance. The results are informative as not only are all of the performance controls highly significant, but their inclusion increase the size of the coefficient on Black from 1.752 to 1.792 significant at the 1% level. The three dummy variables for comparative performance level (poor performance, high performance and very high performance) with middle performance as the omitted category are all significant. Poor performance and very high performance at the 1% level and high performance significant at the 5% level. Interestingly, poor performance and very high performance are associated with the greatest increases in hazard. It could be the fact that while poor performers face increased hazards due to not meeting the required standards, very high performers likely have alternative employment opportunities that result in their increased likelihood to leave the university. This result is further backed up by the two unadjusted continuous variables citations per month and citations per month squared. Both are significant at the 1% level.

Since the Cox Proportional Hazards model does not specify the baseline hazard rate, for that reason I provide it below for black employees and white employees. This graph provides further context to the proportional hazard rates reported in the Cox regressions as well as serves to provide support for the proportional hazards assumption.

Figure 14: Baseline hazard rate



In the second set of models, reported below, I include an additional measure of individual performance. Indicator variables for a promotion within the previous 1 to 12 months, 13 to 24 months and 25 to 36 are included in the model. The reason for the inclusion of these variables is that an employee who has experienced a promotion is more likely to remain in employment for two reasons, (1) the promotion likely indicates higher performance, and (2) the promotion itself likely increases employee motivation in the short term.

Table 8: Cox Proportional Hazard Model set 2

	Full Sample	Entry > t0
Black	1.792***	1.572***
	(0.248)	(0.267)
Coloured	1.114	1.114
	(0.175)	(0.198)
Asian/Indian	0.943	0.996
	(0.183)	(0.213)
Masters	0.719**	0.717**
	(0.097)	(0.114)
PhD	0.617***	0.520***
	(0.09)	(0.093)
Poor performer	2.378***	2.300***
	(0.406)	(0.495)
High performer	1.441**	1.541*
	(0.259)	(0.345)
Top performer	2.835***	2.415***
	(0.510)	(0.570)
Citations / month	1.008***	1.013*
	(0.001)	(0.007)
(Citations / month)^2	0.999***	0.999
	(0.000)	(0.000)
Promotion previous 12	0.578*	0.343***
	(0.167)	(0.143)
Promotion over 12 less 24	1.379	1.277
	(0.349)	(0.455)
Promotion over 24 less 36	1.3	1.268
	(0.396)	(0.526)
Number Failures	502	323
Employees	1767	1085
N	149722	71965

Notes: Tied employment terminations are handled using the Efron approximation. All models include a full set of dummy variables for year of termination, faculty of employment, age in years, seniority, gender, South African nationality, unemployment and GDP. Promotion indicator variables are coded such that they take on the value 1 if an individual experienced a change in seniority in the previous 1-12, 13-24 and 25-36 months respectively. Robust standard errors, clustered by employee id, are reported in parenthesis. Statistical significance indicated by *** (0.01 level), ** (0.05 level) and * (0.10 level).

Under both the full sample and the restricted sample, having experienced a promotion within the previous 12 months is associated with a significantly lower hazard rate (significant at the 10% level in the full sample and the 1% level in the restricted sample). However, even with the inclusion of these additional performance controls, black employees still face dramatically higher hazard ratios. The effect of the other performance measures remain significant and large.

Table 9: Cox proportional hazards result set 3

	Model 7	Model 8	Model 9
Black	3.323*		1.745***
	(2.107)		(0.300)
Coloured	2.032		
	(1.272)		
Asian/Indian	1.808		
	(1.216)		
Masters	0.769*		
	(0.119)		
PhD	0.701*		
	(0.147)		
Poor performer	2.380***	2.353***	2.420***
	(0.407)	(0.403)	(0.414)
High performer	1.449**	1.434**	1.479**
	(0.261)	(0.258)	(0.266)
Top performer	2.827***	2.804***	2.858***
	(0.505)	(0.500)	(0.513)
Citations / month	1.008***	1.008***	1.008***
	(0.001)	(0.001)	(0.001)
(Citations / month)^2	0.999***	0.999***	0.999***
	(0.000)	(0.000)	(0.000)
Promotion 1-12	0.586*	0.592*	0.586*
	(0.169)	(0.171)	(0.168)
Promotion 13-24	1.370	1.379	1.372
	(0.348)	(0.353)	(0.350)
Promotion 25-36	1.302	1.315	1.299
	(0.395)	(0.398)	(0.394)
Educational dissimilarity	1.307	1.984***	1.909***
	(0.451)	(0.465)	(0.451)
Seniority dissimilarity	1.289	1.008	1.182
	(0.838)	(0.654)	(0.771)
Gender dissimilarity	0.513	0.528	0.492
	(0.226)	(0.234)	(0.218)
Racial dissimilarity	0.403	1.428**	1.000
	(0.356)	(0.204)	(0.182)
Age dissimilarity	2.265	2.691	2.393
	(2.722)	(3.213)	(2.854)
Tenure dissimilarity	1.168	1.180	1.154
	(0.146)	(0.147)	(0.143)

Notes: Tied employment terminations are handled using the Efron approximation. All models include a full set of dummy variables for year of termination, faculty of employment, age in years, seniority, gender, South African nationality, unemployment and GDP. Promotion indicator variables are coded such that they take on the value 1 if an individual experienced a change in seniority in the previous 12, 24 and 36 months respectively. Robust standard errors, clustered by employee id, are reported in parenthesis. Statistical significance indicated by *** (0.01 level), ** (0.05 level) and * (0.10 level).

Model result set 3 examines the impact of an alternative set of demographic measures. The variables introduced measure individual employee dissimilarity to their fellow employees within the same faculty at a given point in time. These are dissimilarity in terms of educational attainment, positional seniority, gender, racial classification, age and length of tenure at the university. In the first specification, these additional 5 control variables are included along with educational attainment and individual race. Under that specification, the impact of increased levels of dissimilarity is mixed. Higher levels of dissimilarity are associated with a higher hazard rate for all dissimilarity measures except differences in employee gender and race. However, none of the findings are statistically significant. This is likely the result of a very high degree of correlation between the racial dissimilarity measure and the indicator variable for black race. Similarly, I suspect that correlation between educational attainment and educational dissimilarity may also be high. For this reason I run specification 2 and exclude the indicator variables for an individual's race and the variables for an individual's educational attainment. Under this specification, both educational dissimilarity and racial dissimilarity are significantly (at the 1% and 5% levels respectively) associated with an increased hazard rate. The findings reported above remained robust to additional model specifications (not reported) that included a variable indicating permanent employment instead of contract employment. In these cases, the coefficient on permanent employment was below 1 but did not display statistical significance.

In the third specification I include an indicator variable for black race only (all other races omitted) since I suspect that black employees may experience the workplace differently. Under this specification, the black indicator variable is of similar magnitude to those found in earlier models that excluded the dissimilarity measures.

In my final specification I model interaction effects between age and race, and nationality and race. Inclusion of these interaction terms results in some interesting findings. Age has the effect of decreasing differences in terms of hazard rate between different races. Interestingly, both white employees who hold foreign nationality and black employees who hold foreign nationality face lower levels of departure risk.

Table 10: Interaction model result set

Interactions Model		
	HR	Std. Error
Black	4.225*	(3.134)
Coloured	4.880*	(4.344)
Asian/Indian	1.482	(1.572)
Black*Age	0.982	(0.017)
Asian/Indian*Age	0.989	(0.026)
Coloured*Age	0.962*	(0.023)
South African	0.717	(0.281)
Foreign national black	0.662	(0.301)
Foreign national white	0.904	(0.377)

Notes: Tied employment terminations are handled using the Efron approximation. Model includes a full set of dummy variables for year of termination, educational attainment, and faculty of employment, age in years, seniority, gender, individual performance (as specified in prior models), South African nationality, unemployment and GDP. Promotion indicator variables are coded such that they take on the value 1 if an individual experienced a change in seniority in the previous 12, 24 and 36 months respectively. Robust standard errors, clustered by employee id, are reported in parenthesis. Statistical significance indicated by *** (0.01 level), ** (0.05 level) and * (0.10 level).

For example, a 60 year old black employee faces a hazard rate that is in excess of 40% lower $((0.982)^{60-30} = 0.579)$ than a 30 year old black employee would face.

Discussion and conclusions

Overall, results provide strong evidence that not all groups of employees experience the workplace equivalently. And, that much of one's experience is dependent upon one's race, educational attainment and individual performance. In addition, risk of departure is not constant over an employee's tenure, but instead highest in an employee's early years at the university. This hazard peaks could be the focus of strategic interventions aimed at reducing turnover among employees.

There were two primary issues under investigation in this study. The first considered how individual characteristics influenced turnover and the second examined how turnover is influenced by diverse groups of people working together. Overall, the results of these models are mixed, but it is apparent that race does matter.

The problem is not only that black academics are not being recruited with great enough frequency, but also the fact that these employees are much more likely to experience early termination that their white peers. This has the effect of placing a double burden on university administrators who are attempting to increase levels of diversity among the academic faculty. Further, there is some evidence to suggest that in particular, black academics receive lucrative opportunities from international universities, the private and government sectors.

Overall, the university's experience with diversification has been mixed. A trend analysis of UCT employees revealed promising early gains in the level of diversity which unfortunately were not sustained beyond 2006. Further the detailed trend analysis suggests that diversification efforts while initially quite effective became less so, especially from 2007 onwards.

6.1 Limitations and areas for future investigation

Given the non-experimental setting of this study it is not possible to infer causal relationships which are of particular interest to decision makers. Findings related to the dissimilarity specifications are limited, largely due to very high level of correlation between racial dissimilarity and race itself. Such a study could be replicated using a more diverse university to better understand the effects of diversity in the workplace.

Works Cited

Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories. *Sociological Methodology, 13*, 61-98.

Bender, K. A., & Heywood, J. S. (2006). Job satisfaction of the highly educated: The role of gender, academic tenure, and earnings. *Scottish Journal of Political Economy, 53*(2), 253-279.

Bilmoria, D., & Buch, K. K. (2010). The Search is On: Engendering Faculty Diversity Through More Effective Search and Recruitment. *Change: The Magazine of Higher Learning, 42*(4), 27-32.

Bureau of Labor Statistics. (2014, September 18). *Employee Tenure Summary*. Retrieved from Bureau of Labor Statistics: http://www.bls.gov/news.release/tenure.nr0.htm

Cape Times. (2015, March 23). *UCT deliberately not hiring black profs*. Retrieved from MSN News: http://www.msn.com/en-za/news/featured/uct-deliberately-not-hiring-black-profs/ar-BBiBhnY

Chinomona, R., & Dhurup, M. (2014). The influence of the quality of working life on employee job satisfaction, job commitment and tenure intention in the SME sector in Zimbabwe. *South African Journal of Economic and Management Sciences, 17*(4), 363-378.

Cleves, M., Gutierrez, R., Gould, W., & Marchenko, Y. (2008). *An Introduction to Survival Analysis Using Stata.* (2nd, Ed.) College Station, TX: Stata Press.

Cotton, J. L., & Tuttle, J. M. (1986). Employee Turnover: A Meta-Analysis and Review with Implications for Research. *The Academy of Management Review, 11*(1), 55-70.

Creswell, J. W. (1985). Faculty Research Performance: Lessons from the Sciences and the Social Sciences. *SHE-ERIC Higher Education Report No. 4*.

Ehrenberg, R. G., Kasper, H., & Rees, D. (1990). Faculty Turnover at American Colleges and Universities: Analyses of AAUP Data. *NBER Working Paper Series No. 3239*.

Ghaemi, N. (2009). *A Clinician's Guide to Statistics and Epidemiology in Mental Health: Measuring Truth and Uncertainty.* Cambridge, UK: Cambridge University Press.

Government of South Africa. (2016). *Government Vacancies*. Retrieved 2016, from South African Government: http://www.gov.za/about-government/government-vacancies

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A Meta-Analysis of Antecedents and Correlates of Employee Turnover: Update, Moderator Tests, and Research Implications for the Next Millenium. *Journal of Management, 26*(3), 463-488.

Groundup Staff. (2015, May 14). *The academic staff*. Retrieved from Mail and Gaurdian: http://mg.co.za/article/2015-05-14-uct-and-transformation-the-academic-staff

Groundup Staff. (2015, April 16). *UCT and transformation part two: the students*. Retrieved from GoundUp: http://www.groundup.org.za/article/uct-and-transformation-part-two-students_2845/

Harrigan, M. N. (1999). An Analysis of Faculty Turnover at the University of Wisconsin-Madison. *39th Annual AIR Forum*.

Harzing, A.-W. (2010). *The publish or perish book : your guide to effective and responsible citation analysis.* Melbourne, Australia: Tarma Software Research Pty Ltd.

Haupt, A. (2015, Marcg 20). *It matters who teaches you, Graça*. Retrieved from Mail and Gaurdian: http://mg.co.za/article/2015-03-19-it-matters-who-teaches-you-graca

Hom, P. W., & Kinicki, A. J. (2001). Toward a Greater Understanding of How Dissatisfaction Drives Employee Turnover. *Academy of Management Journal, 44*(5), 975-987.

Jovanovic, B. (1979). Job Matching and the Theory of Turnover. *Journal of Political Economy, 87*(5), 972-990.

Kaplan, E., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association, 53*(282), 457-481.

Kessi, S., Buthelezi, M., Chirikure, S., Chitonge, H., Searle, B., Collis-Buthelezi, V., . . . Mulaudzi, M. (2015, March 17). *University of Cape Town*. Retrieved from It matters who teaches: A response to the UCT Chancellor: http://www.uct.ac.za/dailynews/?id=9025

Kurtulus, F. A. (2011). What types of diversity benefit workers? Empirical evidence of the effects of co-worker dissimilarity on the performance of employees. *Industrial Relations, 50*(4), 678-712.

Lazear, E. P. (1999). Globalization and the Market for Team-Mates. *The Economic Journal, 109*(454), 15-40.

Liu, X. (2012). *Survival Analysis: Models and Applications.* West Sussex, UK: Wiley: Higher Education Press.

Mattox, J. R., & Jinkerson, D. L. (2005). Using survival analysis to demonstrate the effects of training on employee retention. *Evaluation and Program Planning, 28*, 423-430.

Meho, L. I., & Sonnewald, D. H. (2000). Citation ranking versus peer evaluation of senior faculty research performance: A case study of Kurdish scholarship. *Journal of the American Society for Information Science, 51*(2), 123-138.

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. *Journal of the American Society for Information Science and Technology, 58*(13), 2105-2125.

Moed, H. F., Burger, W. J., Frankfort, J., & Raan, A. F. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy, 14*(3), 131-149.

Mumford, K., & Smith, P. N. (2004). Job Tenure in Britain: Employee Characteristics versus Workplace EffectsAuthor. *Economica, 71*(282), 275-298.

Nogcinisa, L. (2015, March 19). *UCT has been resisting transformation for 20yrs* . Retrieved from Politics Web: http://www.politicsweb.co.za/party/uct-has-been-resisting-transformation-for-20yrs--n

O'Connell, S. (2015, January 19). *What UCT's not telling their first-years*. Retrieved from Independent Online: http://www.iol.co.za/capeargus/what-ucts-not-telling-their-first-years-1.1806441

Piercy, F., Giddings, V., Allen, K., Dixon, B., Meszaros, P., & Joest, K. (2005). Improving Campus Climate to Support Faculty Diversity anf Retention: A Pilot Program for New Faculty. *Innovative Higher Education, 30*(1), 53-66.

Price, M. (2014, July 14). *Staff transformation at UCT*. Retrieved from University of Cape Town: http://www.uct.ac.za/dailynews/?id=8752

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata: Volume II: Categorical Responses, Counts, and Survival* (3rd ed.). College Station, TX: Stata Press.

Royalty, A. B. (1998). Job-to-Job and Job-to-Nonemployment Turnover by Gender and Education Level. *Journal of Labor Economics, 16*(2), 392-433.

Royston, P., & Lambert, P. C. (2011). *Flexible Parametric Survival Analusis Using Stata: Beyond the Cox Model.* College Station, TX: Stata Press.

Sackett, P. R., DuBois, C. L., & Noe, A. W. (1991). Tokenism in Performance Evaluation: The Effects of Work Group Representation on Male-Female and White-Black Differences in Performance Ratings. *Journal of Applied Psychology, 76*(2), 262-267.

Singer, J. D., & Willet, J. B. (2003). *Applied Longitudinal Data Analysis.* New York: Oxford University Press.

Sousa-Poza, A., & Sousa-Poza, A. A. (2007). The effect of job satisfaction on labor turnover by gender: An analysis for Switzerland. *The Journal of Socio-Economics*, 895-913.

Tolbert, P. S., Simons, T., Andrews, A., & Rhee, J. (1995). The effects of gender composition in academic departments on faculty turnover. *Industrial and Labor Relations Review, 48*(3), 562-579.

Tsui, A. S., Egan, T. D., & O'Reilly, C. A. (1992). Being Different: Relational Demography and Organizational Attachment. *Administrative Science Quarterly, 37*(4), 549-579.

UCLA. (2014). *Supplemental notes to Applied Survival Analysis Applied Survival Analysis*. Retrieved 12 12, 2015, from Institute for Digital Research and

Education: UCLA: http://www.ats.ucla.edu/stat/examples/asa/test_proportionality.htm

University of Cape Town. (2014). *Our History*. Retrieved from University of Cape Town: http://www.uct.ac.za/about/intro/history/

University of Cape Town. (2014). *Statistics*. Retrieved from University of Cape Town: http://www.uct.ac.za/about/intro/statistics/

University of Cape Town. (2014, December 16). UCT Graduation 2014: Chancellor Graca Machel Speech 2. Cape Town, Western Cape. Retrieved from https://www.youtube.com/watch?v=M9tJ1J7l5dE

University of Cape Town. (2015). *Employment Equity Plan 2015 - 2020*. Retrieved from University of Cape Town: https://www.uct.ac.za/usr/about/transformation/policies/8_EE_Plan_4June_as_of_09h42.pdf

University of Cape Town. (2016). *Cost of Employment (COE) ranges and costing information for permanent academic staff*. Retrieved from University of Cape Town: http://www.hr.uct.ac.za/hr/benefits/remuneration/coe_ranges/academic

Van der Merwe. (2015, November 16). *Interview: UCT Vice-Chancellor Max Price*. Retrieved from Daily Maverick: http://www.dailymaverick.co.za/article/2015-11-16-interview-uct-vice-chancellor-max-price/#.VpR6tPkrKUk

Weinberg, S. L. (2008). Monitoring Faculty Diversity: The Need for a More Granular Approach. *The Journal of Higher Education, 79*(4), 365-387.

Wood, J. L. (2008). Faculty of Color in Academe: What 20 Years of Literature Tells Us. *Journal of Diversity in Higher Education, 1*(3), 139-168.

Appendix

Appendix 1: Termination classification table

Reason reported for termination:		
	Category	Reported reason
1	Better prospects elsewhere	Better prospects; other job opportunity; promotion – higher level elsewhere
2	Career change	Career change; change of work type or interest
3	Family	Child rearing; emigration/relocation; family obligations; health reasons – self or spouse; relocation international; relocation national
4	Completion of Contract	Completion of contract
5	Return to education	Continue education; further studies
6	Death	Death
7	Unhappy at organization	Discomfort with job; discomfort with the way things are done; job security; remuneration package; Unhappy – organizational fit; relationship with manager
8	Retrenchment	Involuntary retrenchment; voluntary retrenchment; restructuring; termination
9	Other	Other; reason not disclosed; unknown

Appendix 2: Employment title classification

Classification of employment seniority	
Seniority	Reported Position
Junior	Assistant lecturer; junior research fellow; lecturer; medical officer; research assistant; research fellow; research officer; teaching assistant
Mid-level	Associate professor; deputy dean; senior lecturer; senior research officer; senior lecturer
Senior	Chair; chief research officer; dean; deputy vice chancellor; director; executive director; honoury; principal research officer; professor; vice chancellor

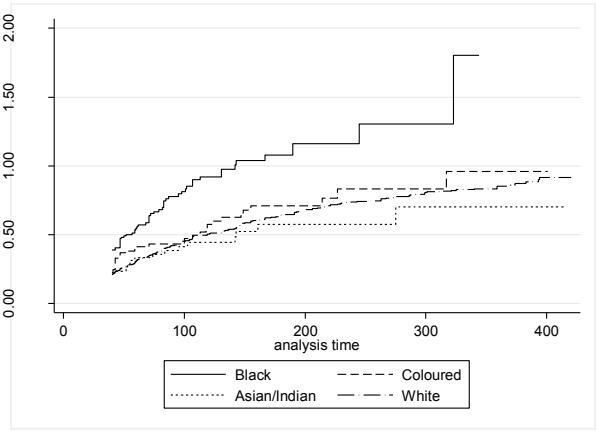
Appendix 3: Observation counts by faculty

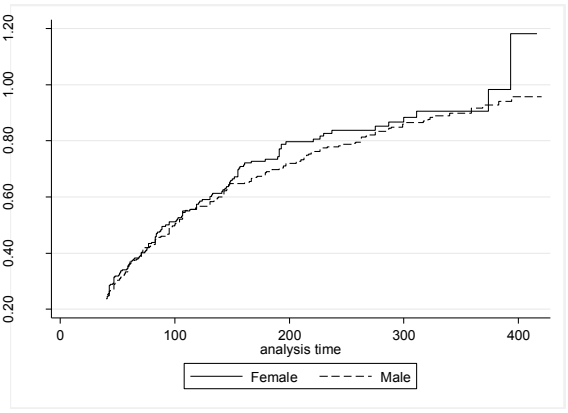
Faculty make up of study Faculty	N (*person-months*)	Percentage
Faculty of Commerce	19872	12.6%
Faculty of Engineering and the Built Environment	21015	13.4%
Faculty of Health Sciences	28574	18.2%
Faculty of Humanities	39277	25.0%
Faculty of Law	8962	5.7%
Faculty of Science	35762	22.7%
Graduate School of Business	3928	2.5%
Total	157390	

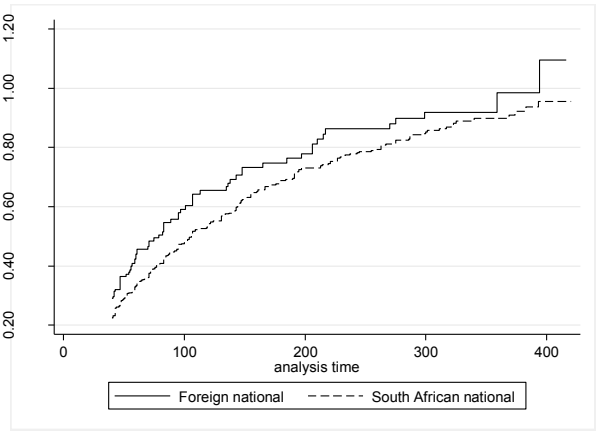
Appendix 4: Classification of educational attainment

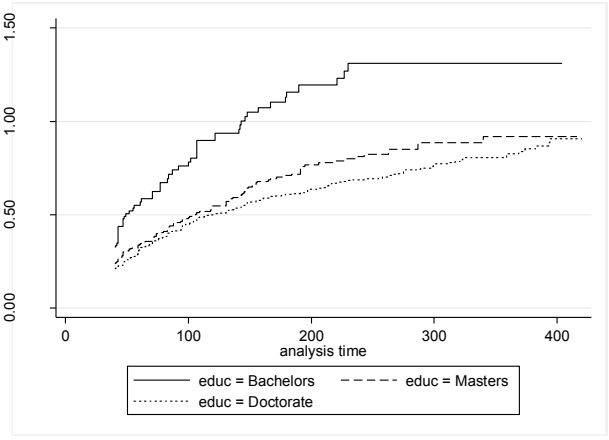
Classification of educational attainment Educational Attainment	Reported Education
Bachelors	General 1st bachelors degree; general academic first bachelor degree; honours degree; MTech degree; other tertiary qualification; PG bachelors degree; PG diploma or certificate; Post-dip diploma; Post graduate bachelors degree; postgraduate diploma/certificate; Prof 1st bachelors degree (4+ yrs); professional first bachelors degree; B.Cur Iet A
Masters	Masters degree; M Dent
Doctorate	Doctoral degree; Professor

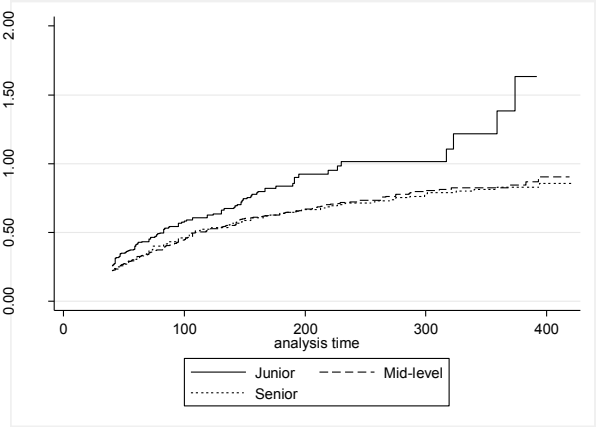
Appendix 5: Nelson-Aalen cumulative hazard curve estimates

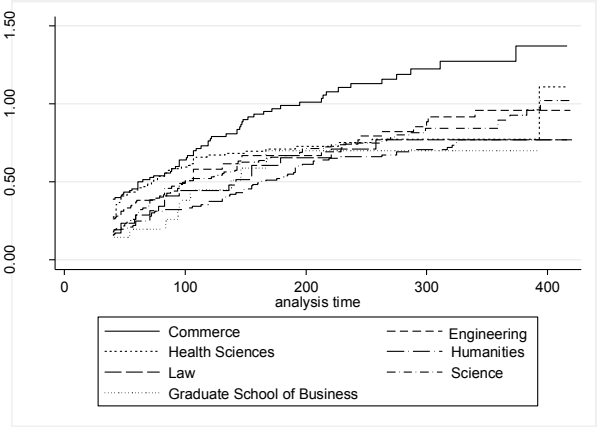




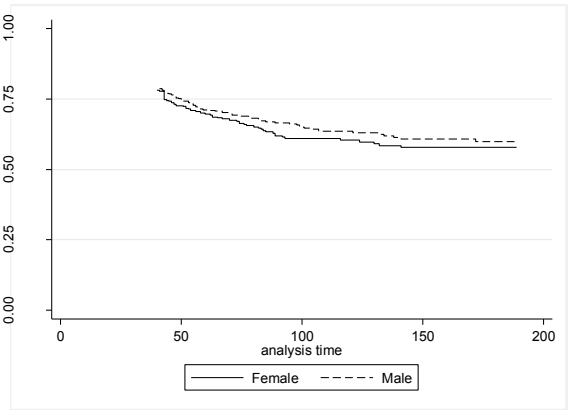


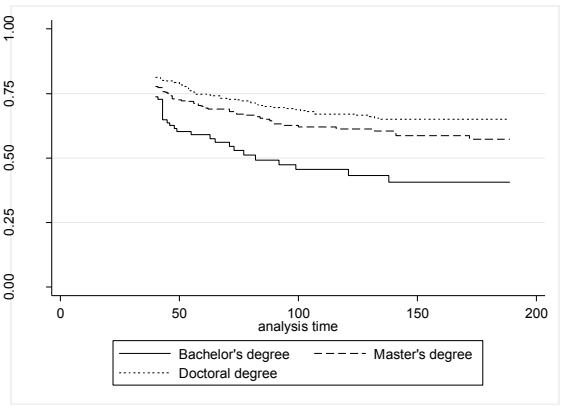


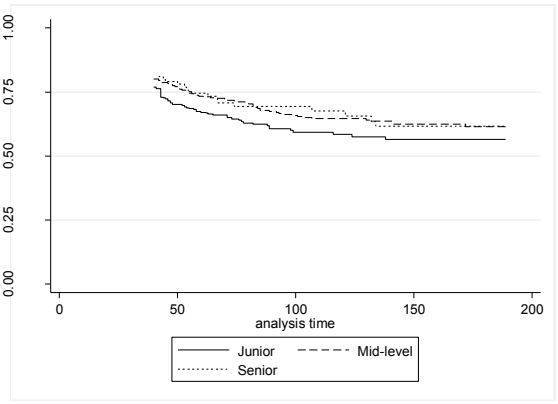


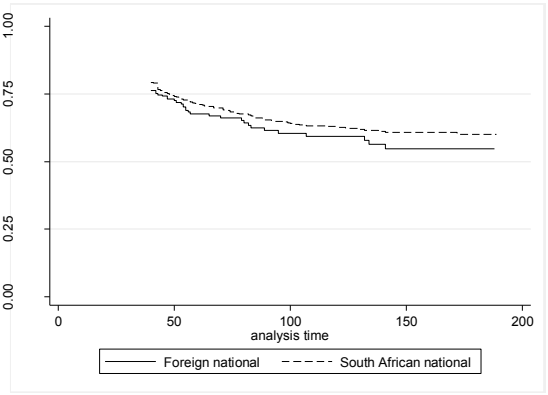


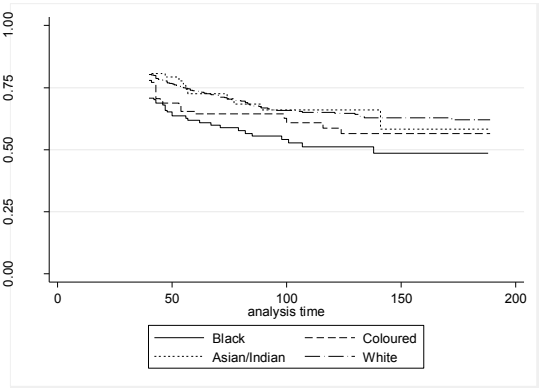
Appendix 6: Restricted sample Kaplan-Meier survival curve estimates

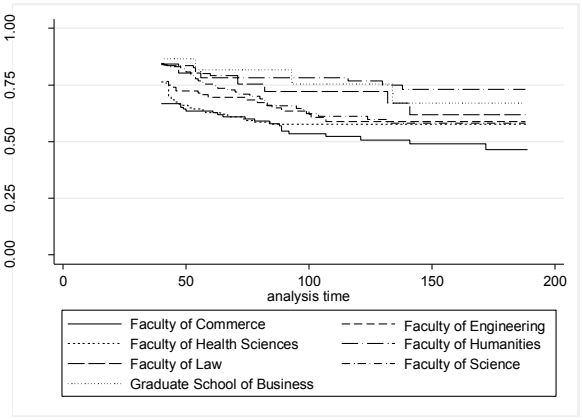












Appendix 7 : Life table

Interval (months)[35]	Total employed at period start	Terminations during period	Lost to censoring	Survival	Standard Error	95% confidence interval
[0-24)	1767	111	125	0.929	0.006	0.92-0.94
[24-48)	1531	130	111	0.848	0.009	0.83-0.86
[48-72)	1290	64	147	0.803	0.010	0.78-0.82
[72-96)	1079	48	71	0.766	0.011	0.74-0.79
[96-120)	960	35	65	0.737	0.012	0.71-0.76
[120-144)	860	33	32	0.708	0.012	0.68-0.73
[144-168)	795	43	73	0.668	0.013	0.64-0.69
[168-192)	679	23	114	0.643	0.013	0.62-0.67
[192-216)	542	29	25	0.608	0.014	0.58-0.64
[216-240)	488	27	25	0.574	0.015	0.54-0.60
[240-264)	439	20	25	0.547	0.015	0.52-0.58
[264-288)	394	24	29	0.512	0.016	0.48-0.54
[288-312)	341	35	50	0.456	0.017	0.42-0.49
[312-336)	256	31	39	0.396	0.018	0.36-0.43
[336-360)	186	20	26	0.350	0.018	0.31-0.39
[360-384)	140	20	26	0.295	0.019	0.26-0.33
[384-408)	94	23	20	0.214	0.020	0.18-0.25
[408-432)	51	16	15	0.135	0.020	0.10-0.18
[432-456)	20	15	5	0.019	0.012	0.00-0.05

35 Following conventional mathematical notation [Brackets] denote inclusion and (parenthesis) denote exclusions.