Economics Working Paper Series

2014/027

Determinants and Timing of Dropping out Decisions: Evidence from the UK FE Sector

Rossella Iraci Capuccinello

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

LUMS home page: http://www.lums.lancs.ac.uk/

Determinants and Timing of Dropping out Decisions: Evidence from the UK FE Sector *

Rossella Iraci Capuccinello †

19 December 2014

Abstract

This paper investigates whether the hazard of dropping out for both male and female students changes over the duration of study. Using duration modelling techniques we find a certain degree of non-monotonic duration dependence for both males and females. However this pattern for female students aiming at high level qualifications is sensitive to attempts to control for unobserved heterogeneity. For these students the extended models show a flattened hazard function, suggesting that the hazard is basically constant over time. For males introducing controls for unobserved heterogeneity does not change the pattern of the duration dependence, suggesting that they might be at higher risk of dropping out during the first semester of their studies.

In addition, we examine variations in drop out hazard patterns for students enrolled on courses which confer different qualification levels. We provide evidence of distinct hazard patterns between students pursuing 'high level' and 'low level' qualifications.

Keywords: Dropout, duration analysis, dropout timing.
JEL Classification: I20, I21, I28.

1 Introduction

The decision to drop out from education can have important consequences for the individual and the society. From an individual point of view dropouts have a higher risk of being unemployed, of living in poverty (Oreopoulos, 2006) and

*The author would like to thank Colin Green, Steve Bradley and Giuseppe Migali for their helpful comments and suggestions. The author is also grateful for comments received from the participants to the 3^{rd} International Workshop in Applied Economics of Education and to the Lancaster University departmental seminar.

†r.iracicapuccinello@lancaster.ac.uk, Department of Economics, Lancaster University Management School, Bailrigg Lancaster LA1 4YX, UK.

of being involved in criminal activities (Lochner and Moretti, 2004). More education improves health outcomes (Silles, 2009) and wages. Moreover, some of the costs related to education will have to be borne even if a student decides to withdraw (fees, part of the study material, etc.). Therefore, in most cases dropping out constitutes a net cost for the individual.

However, the most recents models of dropout behaviour incorporate the possibility of disutility of attending school (Oreopoulos, 2007) and consider dropout behaviour as the result of a learning process where students revise their expectations about the lifetime utility of an extra year of education (Stinebrickner and Stinebrickner, 2012). According to this type of models dropping out from education can be an optimal decision for the students. The drop out behaviour could be determined by the desire to correct a student-course mismatch.

A similar framework can be sketched with regard to the cost of dropouts to society. In fact, when someone drops out of education, society will lose in at least two different ways. On one hand, because of the direct cost of providing education which will have to be borne even in the case that a student decides to withdraw[1]. Secondly, given the relation between dropout, unemployment, health, poverty and inequality, higher dropout rates are likely to generate an increased need for public spending on unemployment and poverty reduction policies.

As part of the Lisbon Strategy the European Council has adopted a set of benchmarks against which to measure the progresses made on the Education and Training 2010 Work Programme. They included among other things the reduction of the percentage of early school-leavers to a maximum of 10 % by 2010 and the achievement of at least 85% of young people with an upper secondary education. Quite naturally some progresses were made in this period, the U.K. early school leavers (age 18-24) proportion was reduced from 18.2% to 17 % in 2008 and the proportion of young people with an upper secondary education went from 76.7% to 78.2 % (European Commission, 2009). However, by 2009 it was clear that many of the objectives set in the Lisbon Strategy were far from being achieved. Therefore, the new Europe 2020 strategy renewed the commitment to achieve such results.

Hence, from a policy point of view it is important to analyse dropout behaviour in order to help policy makers devise possible interventions to reduce it and to target students which are more at risk. In this framework, understanding whether dropout behaviour changes over the school year and whether it does so in an heterogeneous way for different types of students can be a powerful tool in order to ensure that more resources are used in these periods when

[1] At least a part of this costs has to be planned in advance and will have to be borne even if the students drop out, e.g. a teacher salary. A coarse estimate of this cost for 2012/13 is of about 312 millions while the same figure for 2009/10 was 426 millions (Centre for Economic and Social Inclusion, 2015).

students are more in need of support and guidance.

This paper investigates whether the hazard of dropping out from further education sector in England[2] for both male and female students changes over time.

Therefore, the main focus of this paper is to examine not only why students drop out but also when. In this work, we add another dimension to the analysis of the determinants of dropout behaviour by taking into account how the timing of these decisions can affect the understanding of this phenomenon. As can be seen in fig.(1), our raw data shows that dropout probability changes over the school year. Therefore, using a simple discrete choice cross-section analysis would fail to take into account this feature. Hence we employ a useful econometric tool that allows to investigate whether and how the hazard of dropping out changes over time, the duration analysis.
We are also interested in investigating the timing of dropout behaviour for different groups of students. In particular, we carry out separate analysis for students attending a high level course (either academic or vocational) and students aiming at a low level qualification (either academic or vocational). This is important in order to understand whether the timing and type of policies aiming at lowering the dropout rate should be different for these two categories of students. Or whether limited funds could be used more efficiently targeting only students that are more at risk in periods when the risk is higher.

Thinking about the timing of dropping out decisions entails a careful consideration of the differences in terms of outcomes between students dropping out at different moments in time. In other words, we can start thinking about what are the possible advantages or disadvantages of dropping out earlier rather than later.

Estimating higher hazards of dropping out in the first few months after enrolment could show that the decision to drop out is determined mainly by the attempt to correct a student-course mismatch. If this is the case, dropping out would be, in some way, an 'optimal' decision allowing for a process of expectation adjustment (Stinebrickner and Stinebrickner, 2012). Therefore, dropping out earlier would imply a lower waste of time. In this context education could be seen as an experience good and students would not be able to judge its quality before 'consuming' it. Moreover, the opportunity cost of dropping out at an earlier stage is lower than the one of doing it later.

On the other hand, if the risk of dropping out is higher at later periods,

[2]Further education colleges provide both academic courses to help students' progress to higher education at a university or college and career-based vocational and professional courses to help students achieve the skills needed to be successful in the labour market. In the U.K. the FE sector includes: General FE (GFE) and tertiary colleges, Sixth form colleges, Specialist colleges (mainly colleges of agriculture and horticulture and colleges of drama and dance) and Adult education institutions.

we might expect that the student has been able to accumulate some human capital which will be useful in the labour market. Therefore, as advocated by Manski (1989) and Giovagnoli (2005), dropping out of education would not be necessarily a completely negative fact for a student since he or she will anyway have learned something and acquired some useful skills notwithstanding the fact that he or she did not reach a qualification. This argument is supported by the findings of Layard and Psacharopoulos (1974) who show that the economic returns of dropouts are greater than the ones of completers. The authors present evidence against the so called sheepskin effect by which returns are to certificates and not to years of schooling (Jaeger and Page, 1996). The debate between supporters and opponents of the sheepskin effect theory makes evident that analysing the timing of the decision to drop out can tell us something about the reasons and the effects of this choice.
Either way knowing when students are more at risk of dropping out represents an important addition to the existing literature on dropout behaviour since it fosters a more cost-effective implementation of counselling and/or support measures.

From a policy point of view, this paper can help understand the causes of dropout behaviour and devise targeted policies in order to reduce the risk of dropping out. In fact, when implementing an homogeneous model, we find that both female and male students are at higher risk of withdrawal in the first semester of their studies. This could suggest the implementation of better counselling activities and the provision of more information to help them choose the right course. However, when implementing an heterogeneous model, the hazard function for females flattens considerably showing that the risk of dropping out is more or less constant over time. On the other hand, the function for male students continues to show a certain level of non-monotonic duration dependence. This might suggest that male students are more in need of carefully planned pre-enrolment counselling and support services at the beginning of the study year than female students are.

Finally, we also investigate the existence of heterogeneity in the hazard functions for students studying for high or low level qualifications. It is reasonable to expect that students enrolled on a high level qualification will have overall a lower risk of dropping out. We find that when controlling for unobserved heterogeneity female students enrolled on high level qualifications have a risk of dropping out that is constant over time. However, this is not true for the male ones. On the other hand, students of both genders studying for low level qualifications show non-monotonic duration dependence, namely first increasing and then decreasing hazard rate. This suggests that providing these students with better pre-enrolment and counselling services might be an efficient way of reducing drop out rates. It is therefore evident that studying the dropout behaviour with a duration analysis can help enhance the cost-effectiveness of the measures tackling this issue.

2 Literature Review

There is a wide literature on the determinants of dropout behaviour though mainly based on the US and on university studies. On the contrary the literature focusing on further education dropout is quite limited. Several factors affecting the probability of dropping out have been identified and include personal, college and peer characteristics as well as family background and labour market conditions. Females and younger student have a lower probability of dropping out (Johnes and McNabb, 2004; Evans and Schwab, 1995; Smith and Naylor, 2001; Chuang, 1997; Fielding et al., 1998). Students from ethnic minorities are less likely to drop out than white students when other factors such as family or economic background are controlled for (Cameron and Heckman, 2001; Lofstrom, 2007; Bradley and Lenton, 2007). Higher ability students are less likely to drop out (Eckstein and Wolpin, 1999). However Heckman et al. (2006) find that non-cognitive ability has a greater influence on the probability to drop out than cognitive ability.

There has been an intense debate on the effect of socio-economic and family background with some authors finding a significant effect (Bratti, 2007; Sacerdote, 2007) and others a non-significant one (Behrman et al., 2005; Bingley et al., 2008). Another determinant of dropout behaviour for which there is no clear consensus in the literature is the effect of unemployment. In fact, Smith and Naylor (2001) and Bickel and Papagiannis (1988) find a positive relationship while Rees and Mocan (1997) and Peraita (2000) find a negative one and Warren and Lee (2003) and Mocetti (2008) no effect. Evans et al. (1992) find that peer effects affect the students probability of dropping out but when they take into account the endogeneity of peer group formation the effect disappears. On the other hand, Mora and Oreopoulos (2011) find that close peers are indeed affecting their peers probability of dropping out while not-so-close peers have a very limited effect.

There is a smaller literature on the effects of institutional characteristics on dropout behaviour. The student/teacher ratio is often found to be a significant determinant of dropout (McNeal, 1997; Rumberger and Palardy, 2005). While there is no consensus on the effect of school size with some authors finding no effect (Bryk and Thum, 1989; Smith and Naylor, 2001) and some others a positive one (Rumberger and Thomas, 2000).

Even though the educational literature shows various examples of implementation of a duration analysis to study dropout behaviour, most of the papers are targeting dropout from university or they have at their disposal data on dropout that only allow them to identify the year of dropout without specific reference to the month (Aina, 2005; DesJardins et al., 1999; Lassibille and Navarro Gómez, 2008). One important exception is Bradley and Lenton (2007). They analyse the determinants of dropout behaviour in post-compulsory education in the UK implementing a duration analysis and taking into consideration both single risk and competing risks models. Using the Youth cohort study, they find that the hazard of dropping out for students of 17 to 19 years of age is more or less constant over time with the exception of a few spikes at the end of the qualification period. This might suggest that many students drop out just before examination either because they feel unprepared for them or because they find a job just before the end of their studies. However, when they control for unobserved heterogeneity the hazard function flattens showing an hazard which is practically constant over time.

We contribute to this literature by making use of an administrative data set on the whole population of Further Education students in England over a 3 years period (2002-03 to 2004-05). This data has several advantages with respect to the Youth Cohort Study used by Bradley and Lenton (2007) because it records for each student the exact date of enrolment and the exact date of dropout or completion. This feature allows to reduce the measurement error inherent with the calculation of the length of study and the assignment of dropout status using a retrospective diary information about the educational and work activity. Moreover, the use of administrative data on the whole population of FE students allows to widen our population of interest to students aged 16-24 in line with European Council objective of early school leavers reduction.

Like Bradley and Lenton (2007) we control for unobserved heterogeneity. Given the ways in which students are dissimilar and the administrative nature of our data, assuming no unobserved heterogeneity would be unreasonable. It would in fact entail that we are able to control for all possible differences between individuals that may affect the hazard of dropping out. Conversely, students are likely to differ in many unobservable aspects including, for example, motivation and unobserved ability.

3 Econometric Methodology[3]

There are at least two reasons why the timing of dropout behaviour is an important issue. First of all, the dropout decisions could be connected to time-related events. For example, dropout could happen in proximity of exams or as pointed out by Bradley and Lenton (2007) it could happen in periods when

[3]This section draws mainly on Singer and Willett (2003) and on Jenkins (2004)

job vacancies become more often available. Secondly, knowing when students are more at risk of dropping out could help devising time-based interventions. Therefore, understanding how the risk of dropping out changes over time can be a powerful tool for a more efficient use of resources. Students' educational and psychological support services could be strengthened when they are most needed. Moreover, the timing of dropout can suggest which type of intervention could be more effective.

Analysing the determinants of drop out behaviour estimating a cross-section discrete choice model would fail to take into account the fact that the hazard of dropping out might not be constant over-time (Bradley and Lenton, 2007). Therefore, we need to use a different methodological approach, the duration analysis, that allows for the hazard of event occurrence to change over time.

Carrying out a duration analysis in the context of students' education decisions implies determining not only *whether* a student is going to drop out from further education but also *when* is he/she more likely to do so.

We define event occurrence as a change of status from student to dropout. Our data allows to identify the month when a student becomes a dropout. In our analysis the beginning of time is identified by the day of student enrolment (start of qualification). At this moment in time all the individuals in our population are enrolled as students and are therefore at risk of dropping out. However, no one has dropped out yet.

Following Singer and Willett (2003), we choose to use a discrete metric for time since these methods are best suited when the probability of finding individuals that share the same event time is not small. In our data, many students share the same event time since they drop out in the same month. Therefore, the use of a discrete metric is an obvious choice.

In order to deal with questions about events occurrence we have to address the problem of "censoring". Censoring occurs when the researcher cannot know the event time. In fact, fortunately, not all students will drop out during the observation time and some may actually never drop out at all. In both situations,we will not know the event time for those individuals. Quite obviously, the longer the observation period is and the higher the rate of event occurrence, the less censored observations there will be.

In this study, all censoring occurs because of the end of the observation period[4]. This should suffice to guarantee that censoring is not informative[5]. Some students will achieve a qualification before the end of the observation period, as a consequence we will have full information about their event times

[4]The proportion of students transferring to other institutions is about 0.02% and we have not included those students in our analysis

[5]As pointed out by Singer and Willett (2003) the existence of uninformative censoring does not constitute a problem for duration analysis.

and these observations cannot be considered censored in a strict way even if they never dropped out.

A second type of distinction that can be made with respect to different types of censoring is the one between left and right censoring(Singer and Willett, 2003). In our case, defining the start of time at risk as the day of student enrolment guarantees that left censoring will not occur since all the individuals in the data set were not at risk of event occurrence before that day.
The existence of right censoring is due to the fact that event occurrence is not observed either because the event never actually occurs or because the event occurs after the end of the observation period. This type of censoring is present but very limited in our data. However, we have already mentioned that it can be considered an uninformative censoring since it all happens at the end of the observation period.

The estimation of discrete time duration models entails the implementation of common regression techniques for binary choice models to a data set re-organised in the person-period form. Following such re-organisation, each student is present in the data set as many times as the number of months he/she has been at risk of dropping out.

The event we are interested in is dropout and this event can only occur maximum one time for each individual during the observation period. As a consequence, the corresponding dropout variable will take the value 0 in every month but the last one, where it will take the value 1 for all the students which experienced the event. For the students which did not drop out, this variable will take the value 0 in all the periods.

Having explained how our data has been re-organised, we can talk about the hazard function. The conditional probability of event occurrence for individual i in each time period j can be denoted as hazard $h(t_{ij})$. The conditionality implies that the individual has not yet experienced the event in any previous period. More formally, we have a probability density function $P[T_i = j]$ describing the distribution of a random variable, T. This random variable is equal to the period of time in which the individual experiences the event. For example, if an individual i drops out of further education at the third month after enrolment, T_i will be equal to 3. However, we are concerned with the probability of experiencing the event in each period of time conditioned on survival up to that period, therefore we need to take into account the fact that the so-called risk set is changing from one period to another. This is done by employing the conditional probability density function. Therefore, we can express the discrete time hazard as:

$$h(t_{ij}) = P[T_i = j | T_i \geq j] \tag{1}$$

These $h(t_{ij})$ are the main parameters of interest of a duration analysis and together they represent the hazard function. As probabilities, they are

bounded such that $0 \leq h(t_{ij}) \leq 1$. Therefore, our modelling strategy has to be designed in order to avoid hazard values outside this interval. A solution which has often been used for discrete time duration models is applying the complementary log-log transformation. This function is symmetric and allows for comparability with continuous time models. Moreover, for low values of the hazard the logit-hazard and the clog-log hazard are basically indistinguishable (Singer and Willett, 2003).
The hazard function can be graphically represented by plotting the values of the hazard against the time intervals, thus showing a step function.

It is reasonable to think that each individual has a different hazard function. Therefore, as shown by Singer and Willett (1993) we can modify the model by introducing a vector of p observed predictors representing all those characteristics that can help distinguish people with a high risk of dropping out from people with a low risk. Consequently, we will have:

$$h(t_{ij}) = P[T_i = j | T_i \geq j, X_{1ij} = x_{1ij}, X_{2ij} = x_{2ij}, ..., X_{pij} = x_{pij}] \tag{2}$$

where X_{pij} indicates a set of covariates which influence the hazard of dropping out. Therefore, incorporating observed heterogeneity shows that the hazard is dependent from the different values that the predictors will take for each individual.

As previously stated, to allow for comparability with continuous time hazard models, we have decided to apply the complementary log-log transformation. This transformation, in our case, consists in taking the log of the negated logarithm of the probability that students will not drop out in any given period of time. Therefore, more formally:

$$\text{clog-log} = log(-log(1 - h(t_{ij}))) \tag{3}$$

After some preliminary investigations[6], we decided to estimate a non-parametric specification of the hazard function as it is more informative and allows us to analyse what are the different moments in the FE students career where they are more at risk of dropping out. Therefore, the clog-log hazard function will be expressed as:

$$cloglog(h(t_{ij})) = [\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + ... + \alpha_J D_{Jij}] + [\beta_1 X_{1ij} + \beta_2 X_{2ij} + ... + \beta_P X_{Pij}] \tag{4}$$

where the subscript i is referred to the individual, the subscript j to the time period and the subscript P to the predictor. To estimate a model with a non-parametric specification for time we need to make sure that there are events

[6]Following Singer and Willett (2003) we have estimated different models (quadratic, cubic, fourth order and logarithmic) to investigate which model specification provides the better fit without over-parametizing.

taking place in each of the time periods. Consequently, we have followed the suggestion by Jenkins (2004) creating a set of 24 time dummies for the 2002 and 2003 cohort. In fact, since there were very few observations and in some cases no events at all in the months following the 23rd period, we included in the 24th period all of the events taking place afterward. For the same reason, when we analysed the hazard for different qualifications levels we had to reduce the number of periods to 23 for the 2002 cohort and 22 for the 2003 cohort.

We control for several students, teachers and college characteristics. More specifically the X_{Pij} correspond to the following predictors: ethnicity, disability, learning difficulty, student's prior attainment, region, unemployment, college size, proportion of students from disadvantaged background in the college, proportion of teachers with permanent position, teachers' salary, teachers' qualification, student-teacher ratio, proportion of non-white teachers, type of qualification and area of study.

As demonstrated by Prentice and Gloeckler (1978) the equation (4) can be re-expressed as a probability by applying the following transformation:

$$h(t_{ij}) = 1 - exp(-exp(\text{clog-log})) \quad (5)$$

Therefore, we calculated the hazard by applying the transformation in equation (5) to the estimates from our model (equation (4))

$$h(t_{ij}) = 1 - exp(-exp[(\alpha_1 D_{1ij} + ... + \alpha_J D_{Jij}) + (\beta_1 \bar{X}_1 + ... + \beta_P \bar{X}_P)]) \quad (6)$$

where the $\bar{X}_P$ indicate the average value in our data for each predictor.

The set of probabilities estimated through this method is our hazard function. The parameters are estimated via maximum likelihood. The model presented up to this moment is based on the assumption of absence of unobserved heterogeneity. In other words, it is assumed that all differences between observations are captured by the observed covariates. However, this is a strong assumption. Students could, in fact, differ in many unobserved characteristics such as motivation and ability.Therefore, in order to assess the robustness of our estimates we will need to take into account the possible existence of unobserved heterogeneity.

As suggested by Jenkins (2004), a widely used solution to this problem is the introduction of an error term in the c-loglog estimation of the hazard function. In this way, a random intercept model is estimated and the error term is specified in a parametric way. The most commonly used distributions to capture unobserved heterogeneity in discrete duration models are the Gaussian and the Gamma ones.

Ignoring unobserved heterogeneity may lead to an under-estimation of the 'true' duration coefficients when these coefficients are of positive sign and an over-estimation when they are negative (Jenkins, 2004; Van Den Berg, 2001). While it is widely recognised the importance of accounting for unobserved heterogeneity, the practical implementation is often constrained by the computational difficulty. This is the main reason why many discrete time studies choose as in our case to implement the parametric unobserved heterogeneity model with normally distributed error term, since this specification is less computationally demanding. However, as demonstrated by Nicoletti and Rondinelli (2010) mispecifying the distribution of the unobserved heterogeneity does not change the results in a substantive way.

4 Data

We use a large administrative data set (Individualised Learner Record, ILR) that covers the population of students enrolled in the English further education sector to examine students' decisions to drop out from their course of study. We analyse a subset of all the students enrolled in general further education, sixth form or specialist colleges in the years 2002/2003 to 2004-05. We use this data to identify 2 different cohorts of students: the ones enrolled for the first time in 2002 and the ones first enrolled in 2003. Hence, we restrict our sample to full time, full year, non working students, enrolled in 1 or 2 years long courses. We also focus on students aged 16-24 and exclude from our analysis the students which have transferred to other courses as their presence in our data is very limited (0.2 %).

We include a wide set of controls: ethnicity, disability, learning difficulty, student's prior attainment, region, unemployment rate, college size, proportion of students from disadvantaged background in the college, proportion of teachers with permanent position, teachers' salary, teachers' qualification, student-teacher ratio, proportion of non-white teachers, type of qualification and area of study.

Table (1), (2) and (3) show respectively the dropout rate by student characteristics, the dropout rate by gender and some descriptives about college characteristics. In particular, we can notice from table (1) that the share of dropouts by student characteristics does not change substantially between the two cohorts. Table (2) show that there is a lower proportion of dropouts among female students for both cohorts. Furthermore, the dropout rate is quite similar for the 2 cohorts being 11.9 % for students in the 2002 cohort and 11.2 % for the ones in 2003 cohort. The last of our descriptive tables (table 3) shows that the college characteristics for the two cohorts are quite similar.

Since we are implementing a duration model, we transformed the data to obtain a person-period data set. Therefore, we obtained an unbalanced panel where, there is an observation for each interval the student is at risk of dropping out. 24 duration dummies were generated for the 2 cohorts. When analysing the hazard of dropping out for students aiming at different qualification types, we have generated 23 duration dummies for the 2002 cohort and 22 for the 2003 cohort since the last two period did not have enough variability[7]. The start of time is set at the moment of enrolment for each student and we have implemented robustness checks excluding from the sample the students whose course expected duration was too short.

The figure (1) shows the baseline hazard function by gender for the 2002 and the 2003 cohorts. The function shows that the dropout probability in the raw data changes during the year with higher probability of dropping out in the third, fifth and seventh month of each of the two years. However, we can notice a much flatter baseline hazard in the second year of study for both cohorts. Therefore, we will expect a non-monotonic duration dependence in the first twelve months while a somewhat constant pattern in the remaining ones. It also shows that males are more likely to drop out then their female counterparts.

The figure (2) shows the baseline hazard by gender and qualification type for the two cohorts. The baseline hazard for male and female students aiming at high level qualifications shows a lower risk of dropping out than for people aiming at low ones. Moreover, if we look at the function for students with high qualifications there are no substantial differences in the duration pattern for males and females. On the contrary, male students aiming at low level qualifications show a lower probability of dropping out during the second year of their studies than the female ones. Therefore it appears to suggest that female students aiming at high qualifications are just as likely or slightly less likely to dropout than males while the ones aiming at low qualifications are more likely to dropout than males. Moreover, the baseline hazard for female students shows some degree of positive duration dependence in the second year of their studies while the pattern is not so clear for male students.

5 Determinants of dropout

Table(4) and table(5) show the hazard ratios for male and female students of the two cohorts using both the homogeneous and heterogeneous models. As previously stated we control for a wide range of dropout determinants, including both individual and college characteristics and region fixed effects.

[7] Following Jenkins (2004) we have collapsed the last two or three periods into one category.

Since the only case where a likelihood ratio test for unobserved heterogeneity rejects the null hypothesis of no unobserved heterogeneity is the model for female students for the 2003 cohort we will discuss the estimated hazard ratios for the heterogeneous model only in this case. Table (4) and table (5) show hazard ratios therefore we can interpret the coefficients as the proportional effect on the hazard of dropping out of a unit change in the variable of interest.

Given the administrative nature of our data and in the absence of non-cognitive skills test results, we controlled for students' prior attainment as a proxy for their ability[8]. We found that students with no prior qualification are 15% to 41% more likely to drop out than students with a level 2 qualification. Female students with a qualification below level 1 are between 39% and 58% more likely to drop out than female students with a level 2. While for male students the variable is negative but only significant at 10% level for 2003 cohort and non significant for the 2002 one. Having a level 1 prior attainment increases the likelihood of dropping out by 10-13% for male students and by 19-32% for female ones with respect to their fellow students with a level 2. As expected students with a level 3 prior attainment are less likely to drop out than students with a level 2. However, this effect is stronger for females (19 to 33% less likely). Overall, our estimates suggest that the prior attainment is somehow more relevant for female students than for male ones.

By far the variables that have a greater effect on the likelihood to drop out are the ones expressing different qualification types. Our reference category is students aiming at an high level academic qualification. The estimated hazard ratios show that students aiming at all the other qualification types are at least twice as likely to drop out than the ones in the reference category. In particular, students aiming at low level academic qualifications are between 2.8 and 4.5 times more likely to drop out than the ones aiming at high level academic ones. Therefore, we decided to estimate separate duration models for students aiming at high level qualifications and students aiming at low level ones.

Turning to the effect of the disadvantage in the college, we notice that quite surprisingly students in the highest quintile of disadvantage are less likely to drop out than students at the third one. This might be due to the fact that colleges with a higher share of disadvantaged students receive more resources[9] and could concentrate them on preventing dropout behaviour. The other indicators of college quality used in this study are found to have very limited or no effect on students' probability to withdraw from their studies.

[8]Prior attainment is expressed in terms of NVQ levels. As an example: Students with 5 or more GCSE grade A*-C would have a level 2 qualification while students with less than 5 GCSE grade A*-C would hold a level 1.

[9]The indicator of disadvantage in the college is used in the formula to allocate resources to the institutions.

One exception is the ratio of teachers to support staff which shows a positive relationship with dropout suggesting the need for more support staff. College size does not have any effect on student dropout behaviour, even though we have not accounted for the possible endogeneity of this variable.

The students social background is often referred to as one of the major determinants of drop out behaviour. Our estimates show that coming from a deprived area increases the dropout probability in comparison to students living in non-deprived areas by 19.7% to 26.6% for females and by around 15% for males. Other types of disadvantage also increase the likelihood of withdrawing from their studies for female students of about 35% to 53% for females while for males we estimate an increase of 17-18% but only significant at 10% level. However, our other measure of disadvantage included in the estimation for 2002 cohort does not show any effect.

Learning difficulties reduce the likelihood of dropping out probably because of the additional support those students receive from colleges. The same is not true for students with disabilities. Visual or hearing disabilities do not have a significant effect for both male and female students. While Mobility, physical or medical disabilities do increase the probability of dropping out but only for female students. Finally for mental, emotional or behavioural disability the results are not consistent for the 2 cohort though they show a positive effect on dropout. These results suggest that while the support system in place for students with learning disabilities manages to adequately help those students to progress with their studies, students with disabilities are yet at a higher risk of dropping out then students with no disability. There is therefore room for improvements in terms of the disable students support system.

Another commonly analysed determinant of dropout behaviour is the ethnicity. We included dummies for nine different ethnic groups and our result showed heterogeneous effect of being part of a minority on the dropout probability. More specifically, we observe that being Black African, Chinese, Indian or Pakistani reduces the probability of dropping out for both male and female students. However, this effect is stronger for female students. Bangladeshi female students are less likely to drop out than white ones but the result is only significant for the 2002 cohort. On the contrary, their male counterparts are more likely to drop out than white students even though the hazard ratio is only significant for the 2003 cohort. Black Caribbean males are more likely to withdraw from their studies than whites for both cohorts. However, the results for Black Caribbean Females are either not significant or significant at 10% level. Also, other types of black students are significantly more likely to drop out than white students. Overall these findings with the exception of the result for Black-Caribbean are in line with the findings from Bradley and Lenton (2007) and could suggest that minorities do stay on longer in further education probably because they face more labour market discrimination and therefore need to achieve higher qualification levels. Moreover, in many cases we have noticed a higher reduction in dropout probability for females as compared to males. This finding might be due to a sorting effect given the lower participation rates to post-compulsory education of females of these minority groups, whereby the female students that attend post-compulsory education are also the most motivated ones.

We also looked at the effect of the programme area the student is enrolled to and we notice heterogeneous effects of different programme areas on the dropout probability. Enrolling in science courses reduces the probability of dropping out for female students of 8 to 13% with respect to students in humanities while it does not have any statistically significant effect for males. Enrolling in technical or business courses increases the hazard of withdrawing from their studies for both males and female students. In particular we notice that enrolling in technical courses for female students increases this probability of 77% when we account for unobserved heterogeneity. This is perhaps a result of the overrepresentation of male students and teachers in this type of courses. Turning to students enrolled in courses related to the provision of services to people, taking those courses is increasing the probability to drop out as compared to the reference category for female students while it has no effect for males. Lastly, taking basic skills courses decreases the hazard of withdrawal for both males and females. These findings suggest that there is a need for targeting the students support activities of the colleges depending on the students area of study.

Following the literature on dropout behaviour we included in our model a control for the local unemployment rate. However, we found that the unemployment rate for females is not significant and that for males a 1 point increase in the unemployment rate decreases the hazard of dropping out of 1.6% for the 2002 cohort and of 2.9 % for the 2003 cohort. Even though the effect is quite small it is in line with findings from other authors(Bradley and Lenton, 2007) and with the idea that students should invest more in their educations if they have lower outside options.

Finally included region fixed effects. The estimates show that for all cases except for female students of 2003 cohort students living outside the Greater London region are less likely to drop out. As previously mentioned, all our unrestricted models with the exception of the one for females of 2003 cohort show a non statistically significant likelihood ratio test for the existence of normally distributed unobserved heterogeneity. When we look at the results of female students of 2003 cohort we notice that the effect of almost all covariates is reinforced by the introduction of a random intercept.

6 The timing of dropout behaviour

The complementary log-log estimation of the hazard probability with no-frailty, implemented separately for males and females and for the two cohorts, shows noticeable time variation.
In particular, even after controlling for a large set of individual and college characteristics and the use of region fixed effects, for both males and females the hazard rate seems to be much higher in the third, fifth and seventh month for both cohorts (fig.(3) and (4)), showing a non-monotonic duration dependence. This is especially true for the first year of study. However, for the 2002 cohort it also applies to the second year. This result is contrasting with the findings of another study on dropout in further education from Bradley and Lenton (2007) who find a higher risk of dropping out in the last months of the study period, suggesting students might leave before examinations take place to find a job. On the contrary, our finding could indicate that students drop out mainly because of a student-course mismatch. One possible way to reconcile these two views is that students might be reacting to the results of coursework assignments therefore dropping in proximity of these tests.

In fact, the result that students have a higher hazard of dropping out three months after enrolling, could be a sign that their decision to drop out is determined by a mistake about the course choice revealed by the initial assessments results. This line of thought is consistent with the the dropout model used by Stinebrickner and Stinebrickner (2012). Their model is innovative in that it provides an interpretation of dropout behaviour as the result of a learning process whereby students drop out as a consequence of an expectation adjustment. From a policy point of view, this has the implication that improved student counselling practices before enrolment in a further education college could be an effective way to reduce dropout behaviour. Another possible implication is that following the students more intensively in the first semester of the school year could bring better results than an intervention spread equally over the months. However, as previously stated, these results could be biased in the presence of unobserved heterogeneity.

The homogeneous models in fig.(3) and fig.(4) are, in fact estimated assuming the absence of unobserved heterogeneity. In these models, we have, therefore, assumed that we have been able to control for all differences between students that might affect the hazard of dropping out. This assumption is somewhat unrealistic since students differ in many aspects as, for example in motivation or in unobserved ability. Therefore, it is important to assess the possibility that there are unobservable factors influencing our estimates of the hazard function. As a consequence, following Jenkins (2004) we have replicated our model adding a normally distributed random error term.

As previously mentioned there are other ways to take into account the possible existence of unobserved heterogeneity. We have, therefore attempted, at using different methods such as adding a gamma distributed error term or using the mass point approach suggested by Heckman and Singer (1984). Nonetheless, due to computational difficulties related to the size of our dataset the models did not converge. However, there is evidence that using a flexible specification(piecewise) for the hazard function should reduce differences in the coefficients estimated by assuming different distributions for the error term.

We tested for the existence of unobserved heterogeneity using a likelihood ratio test and we failed to reject the null hypothesis of no heterogeneity for the models with males student and for the one with females of the 2002 cohort. On the contrary, when testing for unobserved heterogeneity in the model with female students in 2003 cohort we obtained a p-value of zero, therefore rejecting the hypothesis of no heterogeneity. This is shown also from the plot of the hazard (fig.(3) and fig.(4)) where we can see that only for female students in the 2003 cohort the new baseline hazard looks flatter and is below the non-frailty one. This finding suggests that after controlling for normally distributed unobserved heterogeneity, the hazard of dropping out for female students in the 2003 cohort seems to be more constant over time than clearly non-monotonic. Even though we can still notice some degree of duration dependence, the difference in the risk of dropping out between different periods is now considerably reduced. Therefore, while male student are somewhat more at risk of dropping out in the first semester of study, this is not necessarily true for their female counterparts.
Moreover, this finding for females students is in line with findings of Bradley and Lenton (2007) that after controlling for unobserved heterogeneity find that the hazard of withdrawal is overall constant over time. However, in our case we can notice some gender difference in the way the hazard of dropping out changes over time. This might suggest that male students are more in need of carefully planned pre-enrolment counselling and support services at the beginning of the study year than female students are.

The last part of this paper tries to investigate whether the risk of dropping out changes differently over time for students aiming at an high level qualification and students aiming at a low level one. We do so by estimating the model separately for students enrolled on a high level (academic or vocational) course and students enrolled on a low level (academic or vocational) one. As before we estimate them separately for males and females and we estimate first a homogeneous model and than we assess the robustness to the existence of unobserved heterogeneity. The figures (5), (6), (7) and in (8) show the resulting hazard function for males and females respectively.

As in the general model presented before, the homogeneous model shows some degree of non-monotonic duration dependence for both males and females, enrolled on both high and low level courses. In fact, the step functions shown in fig.(5), (6), (7) and in (8) represent a higher risk of dropping out in the third, fifth and seventh month of the study year for both students aiming at an high level qualification and students aiming at a low level one. This result holds for both males and females but it appears more pronounced for students aiming at low level qualifications and for the fist year of study.

In fact both male and female students aiming at high qualifications show a non-monotonic inverted-U hazard function for the first year of study but it flattens considerably in the second year. This is true for both the 2002 and the 2003 cohort. Moreover, as expected, for both males and females the students enrolled on high level courses are at a lower risk of dropping out at any moment in time.

Following the introduction of a normally distributed error term in the estimated models, we can see that the results present some differences in the hazard function of female students aiming at high qualifications compared to male students aiming at the same type of qualifications.

The first thing we can highlight is that, even though in most of the cases the introduction of the unobserved heterogeneity term reduces the hazard of dropping out, the hazard function for the students enrolled on low level courses still shows a clearly non-monotonic duration dependence. On the other end, the hazard function for students enrolled on high level courses shows substantial differences based on the gender. In fact, while for male students it still shows the same pattern of non-monotonic duration dependence, for females the hazard is constant over time.

From a policy perspective, this finding suggests that students aiming at lower level qualification types should be provided with high quality pre-enrollment counselling and with support services in the first few months of the study year. But also that male students aiming at high level qualifications could be in need of the same type of support.

7 Concluding Remarks

This paper analyses whether and how the risk of dropping out varies with time and whether it varies in different ways for students aiming at high level qualifications and students aiming at low level ones. Therefore, this work adds the time dimension to the study of the determinants of dropout behaviour. This is particularly important as dropout decisions could be linked to particular events such as examinations, start of new courses, timing of job vacancies, ect... Moreover, analysing the timing of dropout behaviour could help achieving a more efficient use of scarce resources by targeting students which are more at risk of dropping out especially in the periods when the risk is higher.

Therefore, we have implemented a discrete time duration analysis. We did so by first assuming the absence of unobserved heterogeneity and subsequently relaxing this assumption. We find that when implementing a homogeneous duration model the risk of dropping out varies over the study year showing a certain degree of non-monotonic duration dependence, namely first increasing and then decreasing. Both male and female students seem to be at a higher risk of dropping out in the first semester of each school year and in particular on the third, fifth and seventh month. This is, perhaps, showing that they drop out after experiencing a mismatch between what they were expecting from their course and how it is in the reality. Alternatively, it is possible that students drop out as a result of receiving negative coursework evaluations.

We also find that the hazard function for males shows a non-monotonic duration dependence even when we implement the estimation taking into account the existence of unobserved heterogeneity. This is not always true for females. In fact, for the 2003 cohort even though we can still notice a slightly higher risk of dropping out in the first few months, now the hazard seems to be approximatively constant over time.

From a policy point of view, the results suggest, first of all, a possibility to target dropout through carefully planned pre-enrolment counselling activities. Secondly, the need for devising specific measures to support students particularly in the initial period of their study. However, finding that after controlling for unobserved heterogeneity the hazard function flattens considerably especially for females, might suggest that male students are more in need of carefully planned pre-enrolment counselling and support services at the beginning of the study year than female students are.

The second research question answered with this work is whether students aiming at high level qualifications and students aiming at low level ones present differences in the way the hazard of dropping out changes over time. We find that when introducing unobserved heterogeneity in the model the hazard function for female students enrolled on high level courses looks much flatter entailing that the hazard is constant over time for this category of students. On the contrary, the hazard function for all students aiming at low level qualifications and for male ones aiming at high level qualifications still shows some degree of negative duration dependence.

Therefore, we can expect that the introduction of policies trying to target the dropout behaviour of students enrolled on low level qualification at the beginning of the study year might help reducing the dropout rate concentrating the use of resources on the people who mostly need them and in the periods when they are most likely to produce effects. The same kind of policies would also benefit male students enrolled in high qualification courses.

Finally, we stress the importance of taking into consideration the timing of dropout decisions as it can help achieving a better understanding of the reasons why students drop out and provide some useful suggestions in terms of policies which might help making a better use of scarce resources.

References

Aina, C. (2005). Parental background and college dropout: evidence from italy. University of Pavia, Department of Economics and Quantitative Methods.

Behrman, J., M. Rosenzweig, and J. Zhang (2005). Does increasing women's schooling raise the schooling of the next generation. reply. *American Economic Review 95*, 1745–1751.

Bickel, R. and G. Papagiannis (1988). Post-high school prospects and district-level dropout rates. *Youth and Society 20*(2), 123–147.

Bingley, P., K. Christensen, and V. M. Jensen (2008, August). The intergenerational transmission of parental schooling and child development – new evidence using danish twins and their children. Working Paper.

Bradley, S. and P. Lenton (2007). Dropping out of post-compulsory education in the U.K: an analysis of determinants and outcomes. *Journal of Population Economics 20*, 299–328.

Bratti, M. (2007). Parents' income and children's school drop-out at 16 in england and wales: evidence from the 1970 british cohort study. *Review of Economic Household 5*, 515–540.

Bryk, A. S. and Y. M. Thum (1989). The effects of high school organization on dropping out: An exploratory investigation. *American Educational Research Journal 26*(3).

Cameron, S. V. and J. J. Heckman (2001). The dynamics of educational attainment for black, hispanic, and white males. *The Journal of Political Economy 109*(2), 455–499.

Centre for Economic and Social Inclusion (2015, February). Achievement and retention in post 16 education. A report for the Local Government Association. Technical report, Local Government Association.

Chuang (1997). High school youths' dropout and re-enrollment behavior. *Economics of Education Review 16*(2), 171–186.

DesJardins, S. L., D. A. Ahlburg, and B. P. McCall (1999). An event history model of student departure. *Economics of Education Review 18*, 375–390.

Eckstein, Z. and K. Wolpin (1999). Why youths drop out of high school: the impact of preferences, opportunities, and abilities the impact of preferences, opportunities, and abilities. *Econometrica 67*(6), 1295–1339.

European Commission (2009). Key competences for a changing world. PROGRESS TOWARDS THE LISBON OBJECTIVES IN EDUCATION AND TRAINING. Analysis of implementation at the European and national levels. Technical report, European Commission.

Evans, W. N., W. E. Oates, and R. M. Schwab (1992). Measuring peer group effects: A study of teenage behavior. *The Journal of Political Economy 100*(5), 966–991.

Evans, W. N. and R. M. Schwab (1995). Finishing high school and starting college: Do catholic schools make a difference? *The Quarterly Journal of Economics 110*(4), 941–974.

Fielding, A., C. R. Belfield, and H. Thomas (1998). The consequences of drop-outs on the cost-effectiveness of 16-19 colleges. *Oxford Review of Education 24*(4), 487–511.

Giovagnoli, P. (2005). Determinants in university desertion and graduation: An application using duration models. *Economica, La Plata L-I*(1-2).

Heckman, J. J. and J. Singer (1984). *Longitudinal analysis of labor market data*. New York: Cambridge University Press.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Jaeger, D. A. and M. E. Page (1996). Degrees matter: New evidence on sheepskin effects in the returns to education. *Review of Economics and Statistics 78*, 733–740.

Jenkins, S. (2004). *Survival analysis*. Mimeo.

Johnes, G. and R. McNabb (2004). Never give up on the good times: Student attrition in the UK. *Oxford Bulletin of Economics and Statistics 66*(1), 23–47.

Lassibille, G. and L. Navarro Gómez (2008). Why do higher education students drop out? evidence from spain. *Education Economics 16*(1), 89–105.

Layard, R. and G. Psacharopoulos (1974). The screening hypothesis and the returns to education. *The Journal of Political Economy 82*(5).

Lochner, L. and E. Moretti (2004). The effect of education on crime: evidence from prison inmates, arrests, and self-reports. *American Economic Review 94*, 155–189.

Lofstrom, M. J. (2007). "Why are Hispanic and African American dropout rates so high?". *The Williams Review 2*, 91–121.

Manski, C. F. (1989). Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon. *Economics of Education Review 8*(4), 305–312.

McNeal, R. B. (1997). High School dropouts: a closer examination of school effects. *Social Science Quarterly 78*(1), 209–222.

Mocetti, S. (2008). Educational choices and the selection process before and after compulsory schooling. *Banca D'Italia Eurosistema 691*. Working Paper.

Mora, T. and P. Oreopoulos (2011). Peer effects on high school aspirations: Evidence from a sample of close and not-so-close friends. *Economics of Education Review 30*, 575–581.

Nicoletti, C. and C. Rondinelli (2010). The (mis)specification of discrete duration models with unobserved heterogeneity: A Monte Carlo study. *Journal of Econometrics 159*, 1–13.

Oreopoulos, P. (2006). The compelling effects of compulsory schooling: Evidence from canada. *The Canadian Journal of Economics 39*(1), 22–52.

Oreopoulos, P. (2007). Do dropouts drop out too soon? wealth, health and happiness from compulsory schooling. *Journal of Public Economics 91*, 2213–2229.

Peraita, C. Pastor, M. (2000). The primary school dropout in Spain: The influence of family background and labor market conditions. *Education Economics 8*(2), 157–168.

Prentice, R. L. and Gloeckler (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics 34*.

Rees, D. I. and H. N. Mocan (1997). Labor market conditions and the high school dropout rate: Evidence from new york state. *Economics of Education Review 16*(2), 103–109.

Rumberger, R. W. and G. J. Palardy (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal 42*(1), 3–42.

Rumberger, R. W. and H. Thomas (2000). The distribution of dropout and turnover rates among urban and suburban high schools. *Sociology of Education 73*(1), 39–67.

Sacerdote, B. (2007). How large are the effects from changes in family environment? a study of korean–american adoptees. *Quarterly Journal of Economics 122*(1), 119–157.

Silles, M. (2009). The causal effect of education on health: Evidence from the united kingdom. *Economics of Education Review 28*(1), 122–128.

Singer, J. and J. Willett (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics 18*(2).

Singer, J. and J. Willett (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press.

Smith, J. P. and R. A. Naylor (2001). Dropping out of university: A statistical analysis of the probability of withdrawal for UK university students. *Journal of the Royal Statistical Society. Series A 164*(2).

Stinebrickner, T. R. and R. Stinebrickner (2012). Learning about academic ability and the college dropout decision. *Journal of Labor Economics 30*(4), 707–748.

Van Den Berg, G. J. (2001). *Handbook of Econometrics*, Volume V, Chapter Duration models: specification, identification and multiple durations. North- Holland, Amsterdam.

Warren, J. R. and J. C. Lee (2003). The impact of adolescent employment on high school dropout: Differences by individual and labor-market characteristics. *Social Science Research 32*, 98–128.

Table 1: Incidence of dropout by characteristics. 2002 and 2003 cohort.

Variable	2002 Cohort				2003 Cohort			
	Females		Males		Females		Males	
	% Dropouts	N. Dropouts	% Dropouts	N. Dropouts	% Dropouts	N. Dropouts	% Dropouts	N. Dropouts
No qualification	11.8	1027	14.7	1,315	11.9	1,055	12.8	1,198
Qualif. < level 1	15.9	86	14.5	116	15.6	91	12.6	124
Qualif. level 1	13.7	2646	14.8	3,176	13.1	3,455	13.3	3,722
Qualif. level 2	8.7	5653	10	5,610	8.2	6,346	9.3	5,934
Qualif. level 3	8.6	441	10.1	367	7	445	8.7	368
Qualif. level 4 or 5	11.1	25	13.6	23	10.1	27	7.8	14
Qualif. unknown	12.3	9587	14.1	10,752	12.3	9,463	13.6	10,698
High academic qualification	5.5	3181	5.8	2,723	5.1	3,182	5.4	2,674
Low academic qualification	15.4	943	16.8	906	15.6	1,156	16.8	977
High vocational qualification	13.1	4437	14.9	5,425	11.9	3,781	12.9	3,956
Low vocational qualification	16.8	3192	14.3	2,215	15.8	3,221	13.1	1,644
Other qualification	12.9	7712	16	10,090	12.7	9,542	14.8	12,807
Disadvantage in the college								
1st quantile	10.6	3431	13	4,046	10.2	3,760	12	4,172
2nd quantile	11.1	3921	13	4,231	11.5	4,765	13.3	5,327
3rd quantile	11.4	4095	12.8	4,320	11.6	4,623	13	4,835
4th quantile	12	4306	13.7	4,650	11.5	4,544	12.4	4,678
5th quantile	10.5	3611	12.1	3,995	8.4	3,083	8.8	2,938
No disadvantage	10.5	12634	12.3	13,909	10	12,815	11.6	14,012
Homeless, traveller or deprived	7	36	11.5	97	11.7	7,861	12.5	7,833
Other disadvantage	-	-	-	-	12.8	206	11.9	213
No learning difficulty	11.1	18338	13	19,888	10.7	19,590	12.1	20,397
Specific: dyslexia or dyscalculia	10.9	509	10.6	704	9.7	568	10.2	795
Multiple	8.3	17	11	40	6.4	25	9.7	52
Other, moderate or severe	9.9	601	10.1	727	10.1	699	9.8	814
No disability	11	18487	12.8	20,307	10.6	19,720	12	20,897
Visual or hearing impairment	9.5	116	12.4	166	11.7	174	11.2	182
Mobility, physical or medical	11.5	439	11.3	386	10.6	465	10.9	424

Variable	2002 Cohort				2003 Cohort			
	Females		Males		Females		Males	
	% Dropouts	N. Dropouts	% Dropouts	N. Dropouts	% Dropouts	N. Dropouts	% Dropouts	N. Dropouts
Mental, emotional or behav. diff.	15.4	70	14.1	74	13.3	90	14.9	114
White	11.4	16079	12.8	16,286	11	17,328	11.8	16,817
Bangladeshi	8.1	173	13.2	390	9	209	12.1	398
Black African	8.5	445	13	707	7.4	508	12	805
Black Caribbean	13.6	602	18.7	798	12.7	655	16	773
Black other or mixed	16.1	587	18	622	14.5	669	16.5	685
Chinese	5.5	120	7.2	183	5.5	103	6.8	147
Indian	6	300	8.5	509	5.2	291	8.6	552
Pakistani	8.1	501	12.2	926	6.9	406	11.6	876
Asian other or mixed	10	222	11.2	348	8.9	228	11.1	387
Other or mixed	10.6	436	13.5	590	10.8	485	13.4	618
Humanities	9.8	8016	11.9	8,115	9.6	7,293	11.3	7,268
Science	9.6	3380	11.8	5,439	9.6	2,989	11.3	4,338
Technical	16.8	449	18	2,964	17.6	402	17.3	2,289
Business	11.5	1136	14.7	1,434	11.3	862	13.4	1,027
Services to people	16.3	4903	15.4	1,509	15	3,722	13.4	1,221
Basic Skills	7.8	194	8.6	262	8.2	255	8	295
East of England	12.7	2062	14.4	2,212	11.3	2,159	13.8	2,481
East Midlands	11	1473	12.6	1,529	11.2	1,627	12.5	1,690
Greater London	11.4	2710	14.6	3,575	10.6	2,843	13	3,497
North East	9.2	1008	10	1,003	8.6	1,024	8.7	965
North West	10.6	3547	11.4	3,371	10.7	3,762	10.8	3,306
South East	11	2640	13.1	3,118	10.4	3,059	11.8	3,437
South West	11.3	1446	13.9	1,645	11.2	1,945	13	2,091
West Midlands	11.7	2325	12.7	2,412	10.8	2,514	11.5	2,526
Yorkshire and the Humber	10.3	1899	12	2,061	10	1,883	10.9	1,916

Table 2: Dropout by gender, 2002 and 2003 cohort.

	Males		Females		Total	
2002 Cohort	%	No.	%	No.	%	No.
Completers	87.2	145,702.0	89.0	157,201.0	88.1	302,903.0
Dropout	12.8	21,359.0	11.0	19,465.0	11.9	40,824.0
Total		167,061.0		176,666.0		343,727.0
2003 Cohort						
Completers	88.1	163,157.0	89.4	175,927.0	88.8	339,084.0
Dropout	11.9	22,058.0	10.6	20,882.0	11.2	42,940.0
Total		185,215.0		196,809.0		382,024.0

Table 3: College characteristics, 2002 and 2003 cohort.

	2002 Cohort		**2003 Cohort**	
Variable	**Mean**	**Stand. Dev.**	**Mean**	**Stand. Dev.**
Proportion of perm. teachers	60.5	23.031	61.547	22.267
Salary of perm. teachers	20.638	5.194	20.947	5.625
Qualification of teachers	60.87	25.453	62.022	25.488
Students to teachers ratio	35.852	37.756	32.862	37.329
Ratio of teach. to support staff	1.477	0.589	1.657	2.351
% of white teachers	91.991	11.158	91.515	11.201
% of non-white teachers	8.009	11.158	8.484	11.201
Size	1636.998	823.902	1726.483	822.02
Size^2	3358574.339	3241092.846	3656458.399	3348247.714

Table 4: Homogeneous and heterogeneous models, 2002 cohort, all covariates.

	Females		*Males*	
Covariates	*Homogeneous* hazard ratio (St. Err.)	*Heterogeneous* hazard ratio (St. Err.)	*Homogeneous* hazard ratio (St. Err.)	*Heterogeneous* hazard ratio (St. Err.)
Prior attainment				
No qualification	1.001	1.002	1.185***	1.185***
	(0.044)	(0.044)	(0.038)	(0.038)
Qualif. < level 1	1.436***	1.440***	1.174	1.175
	(0.124)	(0.131)	(0.110)	(0.110)
Qualif. level 1	1.191***	1.196***	1.132***	1.133***
	(0.027)	(0.028)	(0.026)	(0.026)
Qualif. level 3	0.815***	0.812***	0.869**	0.869**
	(0.056)	(0.056)	(0.061)	(0.061)
Qualif. level 4 or 5	0.842	0.837	0.905	0.905
	(0.265)	(0.272)	(0.297)	(0.302)
Qualif. unknown	1.216***	1.221***	1.245***	1.245***
	(0.020)	(0.020)	(0.020)	(0.020)
Qualification type				
Low academic qualification	2.791***	2.838***	3.266***	3.270***
	(0.043)	(0.043)	(0.043)	(0.043)
High vocational qualification	2.017***	2.033***	2.322***	2.324***
	(0.028)	(0.028)	(0.028)	(0.028)
Low vocational qualification	2.345***	2.370***	2.161***	2.162***
	(0.032)	(0.032)	(0.036)	(0.035)
Other qualification	1.920***	1.932***	2.460***	2.461***
	(0.028)	(0.027)	(0.028)	(0.027)
Disadvantage in the college				
1st quantile	0.897***	0.895***	0.958	0.958
	(0.029)	(0.030)	(0.028)	(0.028)
2nd quantile	0.930***	0.929***	0.952*	0.952*
	(0.026)	(0.027)	(0.026)	(0.026)
4th quantile	1.094***	1.096***	1.087***	1.088***
	(0.026)	(0.027)	(0.025)	(0.026)
5th quantile	0.949*	0.950*	0.918***	0.918***
	(0.029)	(0.030)	(0.028)	(0.028)
Quality of the college				
Proportion of perm. teachers	1.001	1.001	1.001**	1.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Salary of perm. teachers	0.987***	0.987***	0.991***	0.991***
	(0.002)	(0.002)	(0.002)	(0.002)
Qualification of teachers	0.997***	0.997***	0.997***	0.997***
	(0.000)	(0.000)	(0.000)	(0.000)
Students to teachers ratio	1.000	1.000	1.000*	1.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Ratio of teach. to support staff	1.059***	1.060***	1.051***	1.051***
	(0.016)	(0.016)	(0.014)	(0.015)
% of non-white teachers	1.003***	1.003***	1.000	1.000
	(0.001)	(0.001)	(0.001)	(0.001)
size	1.000***	1.000***	1.000***	1.000***
	(0.000)	(0.000)	(0.000)	(0.000)

	Females		*Males*	
Covariates	*Homogeneous* hazard ratio (St. Err.)	*Heterogeneous* hazard ratio (St. Err.)	*Homogeneous* hazard ratio (St. Err.)	*Heterogeneous* hazard ratio (St. Err.)
size^2	1.000***	1.000***	1.000***	1.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Social background				
Homeless, asyl. seeker or traveller	0.824	0.823	0.947	0.947
	(0.198)	(0.201)	(0.130)	(0.129)
Learning difficulty				
Specific: dyslexia or dyscalculia	0.854***	0.850***	0.719***	0.719***
	(0.050)	(0.051)	(0.044)	(0.044)
Multiple	0.684	0.677	0.778	0.778
	(0.274)	(0.282)	(0.186)	(0.187)
Other, moderate or severe	0.804***	0.800***	0.698***	0.698***
	(0.048)	(0.049)	(0.044)	(0.044)
Disability				
Visual or hearing impairment	0.943	0.942	1.080	1.080
	(0.101)	(0.103)	(0.087)	(0.087)
Mobility, physical or medical	1.131**	1.136**	1.020	1.020
	(0.055)	(0.056)	(0.058)	(0.058)
Mental, emotional or behav. Diff.	1.634***	1.647***	1.183	1.184
	(0.134)	(0.140)	(0.140)	(0.140)
Ethnic origin				
Bangladeshi	0.757***	0.752***	1.046	1.046
	(0.085)	(0.087)	(0.060)	(0.060)
Black African	0.635***	0.629***	0.906**	0.906**
	(0.061)	(0.062)	(0.048)	(0.048)
Black Caribbean	1.038	1.037	1.241***	1.241***
	(0.049)	(0.050)	(0.043)	(0.043)
Black other or mixed	1.295***	1.303***	1.265***	1.266***
	(0.048)	(0.049)	(0.047)	(0.047)
Chinese	0.665***	0.660***	0.698***	0.698***
	(0.110)	(0.112)	(0.094)	(0.094)
Indian	0.561***	0.555***	0.664***	0.664***
	(0.064)	(0.065)	(0.050)	(0.051)
Pakistani	0.607***	0.601***	0.923**	0.923**
	(0.058)	(0.058)	(0.039)	(0.039)
Asian other or mixed	0.928	0.924	0.884*	0.884*
	(0.078)	(0.080)	(0.064)	(0.064)
Other or mixed	0.923	0.921	1.067	1.067
	(0.057)	(0.058)	(0.051)	(0.051)
Programme area				
Science	0.921***	0.919***	0.982	0.982
	(0.022)	(0.023)	(0.019)	(0.019)
Technical	1.404***	1.417***	1.271***	1.271***
	(0.052)	(0.054)	(0.024)	(0.024)
Business	1.075**	1.078**	1.342***	1.343***
	(0.034)	(0.035)	(0.031)	(0.031)
Services to people	1.215***	1.221***	1.046	1.046
	(0.021)	(0.021)	(0.031)	(0.031)
Basic Skills	0.642***	0.639***	0.613***	0.613***
	(0.079)	(0.080)	(0.067)	(0.067)

Covariates	*Females* *Homogeneous* hazard ratio (St. Err.)	*Females* *Heterogeneous* hazard ratio (St. Err.)	*Males* *Homogeneous* hazard ratio (St. Err.)	*Males* *Heterogeneous* hazard ratio (St. Err.)
Unemployment rate	0.997	0.996	0.984**	0.984**
	(0.008)	(0.008)	(0.007)	(0.007)
Student region				
East of England	0.983	0.985	0.892***	0.892***
	(0.040)	(0.041)	(0.037)	(0.037)
East Midlands	0.900**	0.900**	0.825***	0.825***
	(0.042)	(0.043)	(0.039)	(0.040)
North East	0.845***	0.844***	0.770***	0.770***
	(0.047)	(0.048)	(0.046)	(0.046)
North West	0.925**	0.926**	0.816***	0.816***
	(0.034)	(0.035)	(0.033)	(0.033)
South East	0.930*	0.930*	0.884***	0.884***
	(0.039)	(0.040)	(0.036)	(0.036)
South West	0.866***	0.865***	0.849***	0.848***
	(0.048)	(0.049)	(0.044)	(0.045)
West Midlands	0.979	0.980	0.877***	0.877***
	(0.036)	(0.036)	(0.033)	(0.033)
Yorkshire and the Humber	0.872***	0.871***	0.826***	0.826***
	(0.039)	(0.040)	(0.037)	(0.037)
lnsig2u		0.248***		0.015***
		(0.251)		(0.837)
Observations	1735268	1735268	1591789	1591789
Log-likelihood	-82587.77	-82588.51	-87461.68	-87461.74
Significance levels : $*$: 10%	$**$: 5%	$***$: 1%		

Table 5: Homogeneous and heterogeneous models, 2003 cohort, all covariates.

	Females		*Males*	
Covariates	*Homogeneous* hazard ratio (St. Err.)	*Heterogeneous* hazard ratio (St. Err.)	*Homogeneous* hazard ratio (St. Err.)	*Heterogeneous* hazard ratio (St. Err.)
Prior attainment				
No qualification	1.272***	1.414***	1.153***	1.156***
	(0.046)	(0.062)	(0.043)	(0.043)
Qualif. < level 1	1.393**	1.580**	0.784*	0.784*
	(0.164)	(0.228)	(0.141)	(0.141)
Qualif. level 1	1.228***	1.326***	1.103***	1.104***
	(0.029)	(0.04)	(0.029)	(0.029)
Qualif. level 3	0.728***	0.672***	0.911	0.910
	(0.066)	(0.085)	(0.068)	(0.069)
Qualif. level 4 or 5	1.176	1.245	0.584	0.582
	(0.259)	(0.357)	(0.411)	(0.412)
Qualif. unknown	1.260***	1.381***	1.184***	1.188***
	(0.023)	(0.032)	(0.023)	(0.023)
Qualification type				
Low academic qualification	3.120***	4.531***	3.471***	3.518***
	(0.047)	(0.077)	(0.049)	(0.050)
High vocational qualification	2.080***	2.537***	2.267***	2.281***
	(0.034)	(0.05)	(0.034)	(0.034)
Low vocational qualification	2.203***	2.669***	2.055***	2.063***
	(0.039)	(0.058)	(0.047)	(0.047)
Other qualification	2.158***	2.637***	2.502***	2.519***
	(0.031)	(0.047)	(0.032)	(0.031)
Disadvantage in the college				
1st quantile	0.989	0.982	0.988	0.987
	(0.035)	(0.048)	(0.034)	(0.035)
2nd quantile	1.013	1.025	1.026	1.026
	(0.031)	(0.042)	(0.029)	(0.030)
4th quantile	1.099***	1.142***	0.914***	0.913***
	(0.032)	(0.043)	(0.031)	(0.032)
5th quantile	0.907***	0.880***	0.794***	0.792***
	(0.037)	(0.049)	(0.036)	(0.036)
Quality of the college				
Proportion of perm. teachers	0.999	0.999	1.000	1.000
	(0.001)	(0.001)	(0.000)	(0.000)
Salary of perm. teachers	0.985***	0.980***	0.987***	0.987***
	(0.002)	(0.003)	(0.002)	(0.002)
Qualification of teachers	0.999**	0.999**	0.999***	0.999***
	(0.000)	(0.001)	(0.000)	(0.000)
Students to teachers ratio	1.000**	1.001*	1.001***	1.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Ratio of teach. to support staff	1.098***	1.131***	1.091***	1.092***
	(0.017)	(0.024)	(0.017)	(0.017)
% of non-white teachers	1.004***	1.006***	1.003***	1.003**
	(0.001)	(0.002)	(0.001)	(0.001)
size	1.000***	1.000***	1.000***	1.000***
	(0.000)	(0.000)	(0.000)	(0.000)

Covariates	*Females* Homogeneous hazard ratio (St. Err.)	*Females* Heterogeneous hazard ratio (St. Err.)	*Males* Homogeneous hazard ratio (St. Err.)	*Males* Heterogeneous hazard ratio (St. Err.)
size^2	1.000***	1.000***	1.000***	1.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Social background				
Student from deprived area	1.197***	1.266***	1.153***	1.155***
	(0.023)	(0.031)	(0.023)	(0.023)
Other disadvantage	1.354***	1.534***	1.178*	1.182*
	(0.094)	(0.132)	(0.098)	(0.099)
Learning difficulty				
Specific: dyslexia or dyscalculia	0.762***	0.686***	0.705***	0.702***
	(0.057)	(0.077)	(0.051)	(0.051)
Multiple	0.501**	0.374***	0.803	0.799
	(0.285)	(0.373)	(0.198)	(0.203)
Other, moderate or severe	0.883**	0.849**	0.760***	0.757***
	(0.055)	(0.074)	(0.051)	(0.052)
Disability				
Visual or hearing impairment	1.112	1.162	1.002	1.002
	(0.101)	(0.138)	(0.102)	(0.103)
Mobility, physical or medical	1.127*	1.201**	0.908	0.908
	(0.062)	(0.084)	(0.071)	(0.072)
Mental, emotional or behav. diff.	1.202	1.344	1.407**	1.416**
	(0.15)	(0.204)	(0.141)	(0.142)
Ethnic origin				
Bangladeshi	0.898	0.852	1.159**	1.161**
	(0.095)	(0.126)	(0.073)	(0.073)
Black African	0.609***	0.510***	0.898**	0.896**
	(0.066)	(0.087)	(0.053)	(0.054)
Black Caribbean	1.105*	1.137*	1.215***	1.218***
	(0.055)	(0.076)	(0.052)	(0.053)
Black other or mixed	1.250***	1.356***	1.291***	1.295***
	(0.053)	(0.076)	(0.053)	(0.054)
Chinese	0.513***	0.406***	0.666***	0.662***
	(0.138)	(0.175)	(0.11)	(0.111)
Indian	0.569***	0.467***	0.813***	0.810***
	(0.076)	(0.098)	(0.056)	(0.057)
Pakistani	0.592***	0.504***	0.966	0.966
	(0.068)	(0.088)	(0.047)	(0.047)
Asian other or mixed	0.857*	0.814*	0.985	0.985
	(0.091)	(0.119)	(0.068)	(0.068)
Other or mixed	0.983	0.993	1.069	1.071
	(0.064)	(0.085)	(0.056)	(0.057)
Programme area				
Science	0.899***	0.867***	0.966	0.965
	(0.025)	(0.034)	(0.022)	(0.022)
Technical	1.458***	1.770***	1.260***	1.265***
	(0.061)	(0.089)	(0.029)	(0.029)
Business	1.111**	1.172***	1.307***	1.313***
	(0.041)	(0.056)	(0.038)	(0.038)
Services to people	1.163***	1.255***	0.944	0.944
	(0.024)	(0.034)	(0.036)	(0.037)

	Females		*Males*	
Covariates	*Homogeneous* hazard ratio (St. Err.)	*Heterogeneous* hazard ratio (St. Err.)	*Homogeneous* hazard ratio (St. Err.)	*Heterogeneous* hazard ratio (St. Err.)
Basic Skills	0.629***	0.552***	0.605***	0.602***
	(0.072)	(0.094)	(0.066)	(0.066)
Unemployment rate	0.994	0.993	0.971***	0.971***
	(0.011)	(0.014)	(0.011)	(0.011)
Student region				
East of England	1.011	1.028	0.905*	0.905*
	(0.056)	(0.074)	(0.052)	(0.052)
East Midlands	1.135**	1.206**	0.964	0.963
	(0.054)	(0.073)	(0.051)	(0.051)
North East	1.111*	1.153*	0.955	0.955
	(0.059)	(0.081)	(0.06)	(0.061)
North West	1.147***	1.211***	0.914**	0.913**
	(0.047)	(0.063)	(0.045)	(0.046)
South East	1.067	1.104	0.839***	0.838***
	(0.052)	(0.069)	(0.049)	(0.049)
South West	1.123**	1.167*	0.886**	0.885**
	(0.059)	(0.079)	(0.056)	(0.057)
West Midlands	1.149***	1.208***	0.981	0.981
	(0.045)	(0.06)	(0.043)	(0.044)
Yorkshire and the Humber	1.050	1.071	0.988	0.989
	(0.052)	(0.069)	(0.048)	(0.049)
lnsig2u		3.453***		0.162***
		(0.112)		(0.309)
Observations	1385857	1385857	1275353	1275353
Log-likelihood	-62965.42	-62952.75	-65805.76	-65806.29
Significance levels : $*$: 10%	$**$: 5%	$***$: 1%		

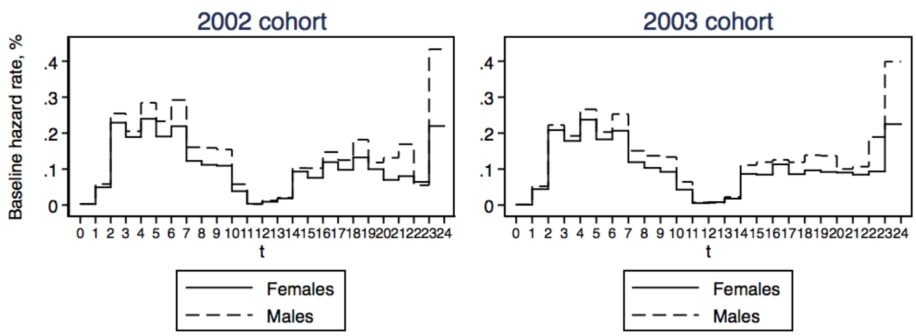

Figure 1: Baseline Hazard by gender, 2002 and 2003 cohorts.

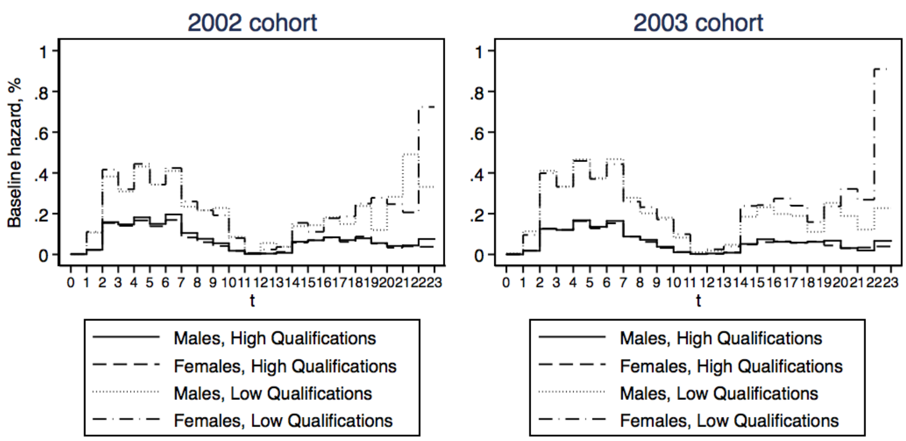

Figure 2: Baseline Hazard by gender and qualification type, 2002 and 2003 cohorts.

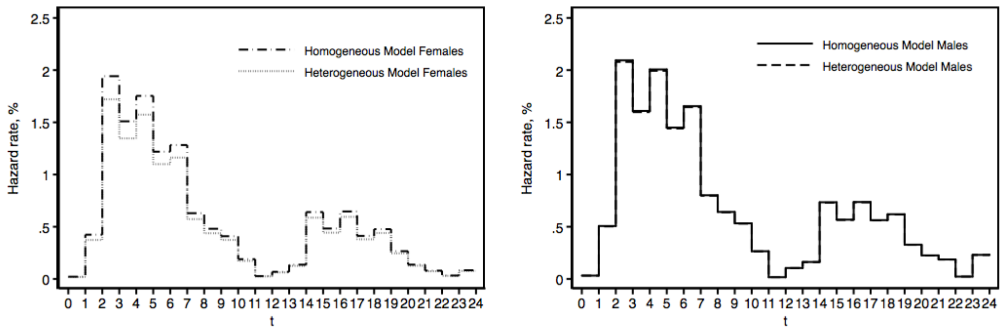

Figure 3: Hazard function by gender, 2002 cohort. Homogeneous and heterogeneous models (Normal error).

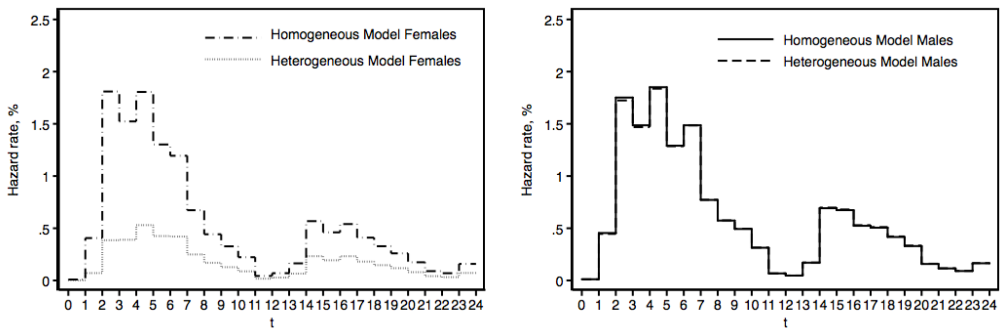

Figure 4: Hazard function by gender, 2003 cohort. Homogeneous and heterogeneous models (Normal error).

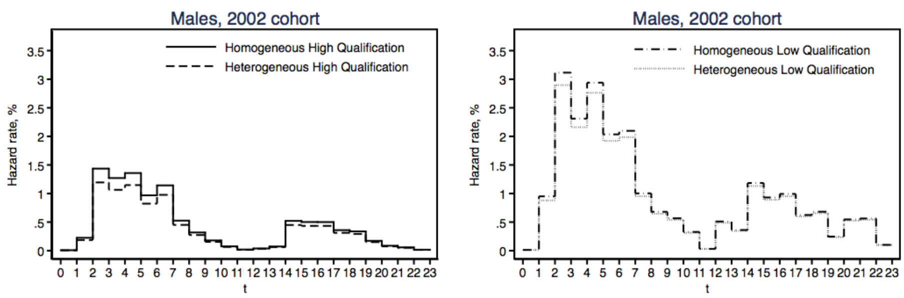

Figure 5: Hazard rate, Males, High vs. Low Qualification Type, 2002 cohort. Homogeneous and heterogeneous models (Normal error).

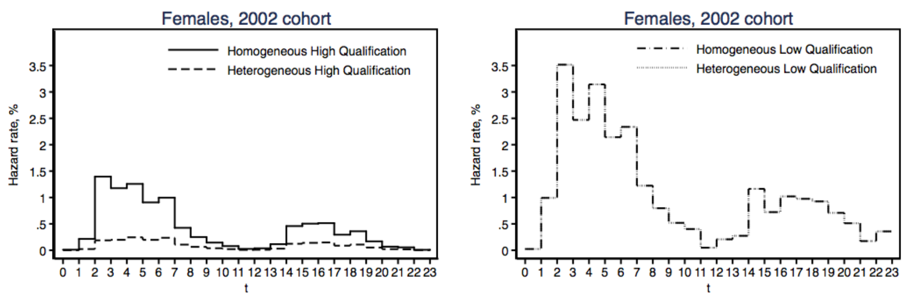

Figure 6: Hazard rate, Females, High vs. Low Qualification Type, 2002 cohort. Homogeneous and heterogeneous models (Normal error).

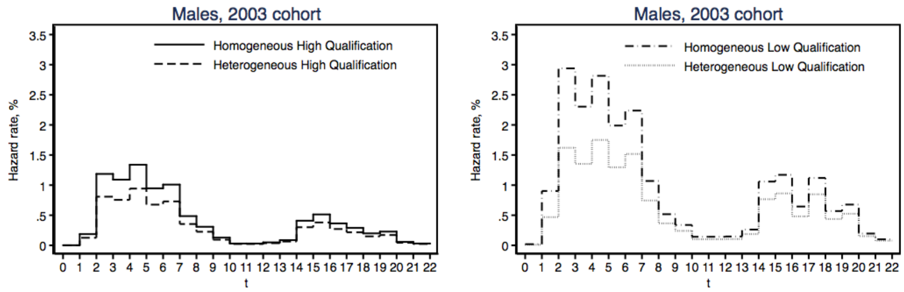

Figure 7: Hazard rate, Males, High vs. Low Qualification Type, 2003 cohort. Homogeneous and heterogeneous models (Normal error).

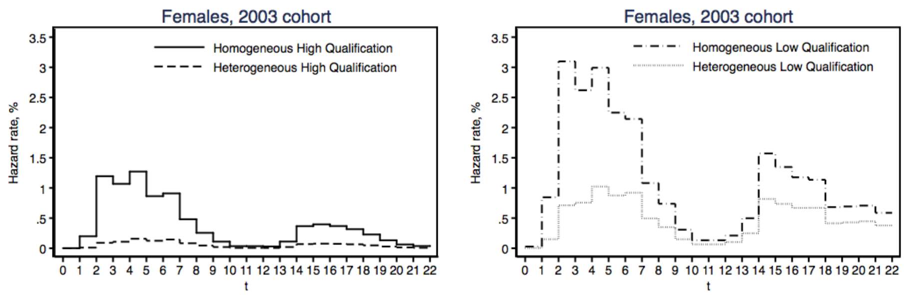

Figure 8: Hazard rate, Females, High vs. Low Qualification Type, 2003 cohort. Homogeneous and heterogeneous models (Normal error).