Labour Market Outcomes of Older Versus Younger Apprentices: A Comparison of Earnings Differentials

Steven McIntosh and Damon Morris

Department of Economics, University of Sheffield

May 2018

Abstract:

Over the last decade, there has been a large increase in the number of individuals completing an apprenticeship, with the growth in numbers particularly strong amongst those age 25+ when starting their apprenticeship. This paper analyses the wage differentials earned by those who complete their apprenticeship, relative to those who start an apprenticeship at the same level but do not complete. The differentials are estimated using a difference-in-differences framework, thus controlling for differences between groups captured by pre-apprenticeship earnings. The results consistently show that individuals who began their apprenticeship when aged 19-24 receive a larger increase in their daily earnings post-completion, relative to non-achievers, than individuals who began their apprenticeship when aged 25+. Subsequent analysis shows that for women with Intermediate and Advanced Apprenticeships, and for men at the Intermediate level only, this difference between age groups is mostly due to the older apprentices receiving a smaller differential within the same framework. For males with Advanced Apprenticeships, the difference between age groups is mostly due to the older apprentices training in areas with lower differentials.

Keywords: Apprenticeships; earnings differentials

JEL Codes: J24; I28

Acknowledgements:

This paper forms part of the work programme on apprenticeships under the Centre for Vocational Education Research (funded by the Department for Education). The administrative data used were kindly matched and provided by the Department for Education. All interpretations of the results remain our own. We are grateful for helpful comments provided by seminar participants at the University of Sheffield, and by CVER colleagues.

Executive Summary

- In the previous 10-15 years, and particularly since 2010, there has been a large increase in the number of apprentices in England, with the growth particularly marked amongst those aged 25+ when starting their apprenticeship. We therefore analyse whether the earnings differentials received by those who complete an apprenticeship are as high now as when estimated in earlier literature at a time of lower apprenticeship participation. In addition, the key research question is whether the differentials received by older apprentices (aged 25+) are as large as those for apprentices in the 19-24 year old age group.
- We use administrative data from the ILR matched to HMRC tax and benefit records. The sample used is all individuals who began an Intermediate (Level 2) or Advanced (Level 3) Apprenticeship between 2004-2013, who fall into either the 19-24 or 25+ age groups (so 16-18 year old apprentices are not included) and for whom earnings are observed in all years considered before and after their apprenticeship.
- The analysis was conducted using a difference-in-differences (DiD) methodology. The treatment group was all individuals who completed an apprenticeship at a given level, while the control group was all individuals who started an apprenticeship at the same level, but did not complete for some reason. The DiD methodology, by controlling for pre-apprenticeship earnings, ensures that any differences in unobserved characteristics between completers and non-completers that influence earnings are differenced out, to the extent that such characteristics remain constant over time. The method also controls for anything else that changes over time that might affect earnings, to the extent that such changes affect treatment and control groups equally. The DiD effects are estimated in a regression framework, controlling for duration of apprenticeship, age completed the apprenticeship, ethnicity, and current tax year dummies for observed earnings. Separate equations are estimated by gender and apprenticeship level. Different specifications are also considered using 1, 2 and 3 years of earnings data either side of the apprenticeship.
- The results show that for every pair of estimates (12 pairs: 2 genders x 2 levels x 3 data windows), the DiD daily earnings differential received by 19-24 year old apprentices is greater than that for apprentices aged 25+. Hence the increase in earnings following completion of an apprenticeship, relative to the change in earnings for non-achievers, is always larger for the younger apprentice group. In most cases, the differential is around twice as large.
- Various robustness checks on the results are made: (i) relaxing the requirement that individuals are observed in <u>all</u> years either side of their apprenticeship; (ii) considering only 25-29 year old apprentices in the older age group, so that they are more similar in age to the younger age group; and (iii) estimating the treatment effects by propensity score matching rather than DiD. In each case, the pattern of results remains the same, with the younger apprentices receiving higher earnings differentials following completion of their apprenticeship. Further checks show, in almost all cases, no significant treatment effect when a 'fake' placebo treatment before the actual apprenticeship is used, as would be expected. Finally, using a comparison group for the younger age group who are one level below the treated apprentices in terms of their highest qualification (rather than using the

non-achievers control group) generally produces similar, slightly larger estimated treatment effects. There is therefore no suggestion of any upward bias from using the non-achievers control group.

- The final research question considers to what extent the older apprentices receive a lower earnings differential because they earn a smaller differential than younger apprentices within the same Apprenticeship Framework, and to what extent it is due to older apprentices typically undertaking apprenticeships in frameworks that earn lower differentials, on average. The results show that for women with Intermediate and Advanced Apprenticeships, and for men at the Intermediate level only, this difference between age groups is mostly due to the older apprentices receiving a smaller differential within the same framework. For males with Advanced Apprenticeships, the difference between age groups is mostly due to the older apprentices training in areas with lower differentials.
- It is therefore important to consider quality of apprenticeships undertaken by older apprentices, in terms of duration, training intensity, etc, to determine why they are valued at a lower level in the labour market.

1. Introduction

The wage differentials associated with qualification attainment are an important source of information about the economic value placed on such qualifications by the labour market, which is of interest to policy-makers making decisions about allocation of funding, and to young people making decisions about their education. As such, a large literature exists on the so-called 'returns to education' (Card, 1999; Harmon and Oosterbeek, 2000). This paper adds to that literature, focussing on a particular form of learning, apprenticeships, obtained in England. The analysis uses difference-in-differences methodology, comparing a treatment group of apprenticeship completers to a control group of apprenticeship non-achievers, both observed before and after they began their apprenticeship. Such methodology allows us to make more causal inference as to the impact of apprenticeship completion on earnings, compared to a standard cross-sectional regression approach. The analysis also explicitly compares the differentials obtained by older apprentices (aged 25+) compared to those of younger apprentices (aged 19-24), which is the first time this question has been studied in the literature.

Since 2010, there has been a huge increase in the number of people undertaking apprenticeships in England. In the 2005/6 academic year, there were 172,600 apprenticeships starts in England, counting all age groups and all levels. By 2010/11, this had increased to 453,000, with numbers peaking at 515,000 in 2012/13. A target was set in 2015 by the incoming government of three million more new starts by 2020. This large change was mostly achieved through a significant increase in the participation of those aged 25 and over in the apprenticeship programme. In 2005/6, there were 74,200 apprenticeship starts by people aged 19-24, compared to just 300 by individuals aged 25+. By 2009/10 these numbers had increased to 113,800 and 49,100, respectively, followed by very large increases in 2010/11 to 143,300 and 182,100 respectively. As these figures show, it is the over 25 group who are responsible for much of this large growth in apprenticeship starts. Since then, there have been further, if smaller, increases in apprenticeship starts by such older workers, reaching a peak of 230,300 in 2012/13, representing 45% of all apprenticeships starts that year.[1]

Previous research has shown that, on average, apprenticeships are associated with large wage or earnings differentials, for example McIntosh (2007), BIS (2011), Conlon *et al.* (2011), National Audit Office (2012). However, given the large recent increases in the number of apprenticeships undertaken, this is a pertinent time to revisit this evidence. What held in a period of lower apprenticeship participation will not necessarily still hold when many more apprentices are appearing on the labour market.

In addition, no previous research has examined in detail how such differentials vary by age group, and whether those aged 25+ see their earnings increase by as much post apprenticeship as those aged 19-24. Again, given the large increase in the number of such older apprentices, this has become an important research question to address. As well as the increased supply of labour from apprenticeships, it is also the case that almost all (91%) of apprentices aged 25+ already worked for their employer before their apprenticeship, compared to 61% of apprentices aged 19-24 and just 36% of apprentices aged 16-18. This at least allows for the possibility that there are differences in

[1] All apprenticeship starts also include 16-18 year olds. Source for all numbers Skills Funding Agency FE Data Library: Apprenticeships (https://data.gov.uk/dataset/fe-data-library-apprenticeships).

the type of training received by different age groups, and so justifies looking into any associated differences in labour market value.

The final contribution of this paper is to make use of administrative data, recently made available in the UK, and described in detail in a subsequent section. Given the need to consider the experiences of recent apprentices in the period of apprenticeship expansion detailed above, survey data sets covering a sample of the working-age population are very unlikely to contain sufficient numbers of recent apprentices. We therefore make use of administrative data that contain the population of funded learners in English Further Education, including apprentices, matched to tax records so that earnings are observed.

The next section briefly reviews research in this area, followed by sections describing the data and methodology used. Section 5 describes the results obtained, while a final section offers discussion and conclusions.

2. Literature Review

There is a large literature examining the wage returns to education, which will not be reviewed here. Within this literature, a far smaller number of papers have focussed explicitly on apprenticeships. Early papers in the UK tended to use survey data sets, in particular the Labour Force Survey (LFS). An example is McIntosh (2007), which used LFS data from 2004/5. The analysis is disaggregated by level of apprenticeship (Level 2, Intermediate, and Level 3, Advanced) and by gender of apprentice. In each case, the weekly earnings of those full-time workers with an apprenticeship are compared to those of an appropriate control group, comprising those whose highest qualifications are the ones typically held by apprentices *before* their apprenticeship. The control group therefore provide an estimate of the counterfactual earnings, that those who complete an apprenticeship would have continued to receive had they not undertaken their training.

The results in McIntosh (2007) reveal substantial weekly earnings differentials between full-time individuals who complete an apprenticeship, and the control group, of around 18% at Level 3 and 16% at Level 2. Differentiating by gender reveals higher differentials associated with apprenticeships for males, with a differential of around 14% at Level 3 for females, and no significant differential at all observed at Level 2. There is also variation in differentials according to the sector in which the individual works,[2] with larger differentials obtained in more 'traditional' apprenticeship sectors such as construction and engineering. Unusually in this literature, McIntosh (2007) also explicitly compares these earnings differentials to estimated costs of providing apprenticeship training,[3] producing large positive net present value estimates.

Other papers in the policy area using LFS data and following a similar methodology have produced very similar results, for different periods (for example see BIS, 2011; Conlon *et al.*, 2011; National Audit Office, 2012). More recently, some studies have made use of the administrative data used in this paper. Most do not focus exclusively on apprenticeships, and so do not offer the disaggregations

[2] No information is available in the LFS recording the sector in which the apprenticeship was actually undertaken, and so the assumption of the paper was that the current sector of employment reflects the sector of apprenticeship.

[3] The costs of apprenticeship training are taken from Hogarth and Hasluck (2003).

of results presented here, nevertheless studies have included apprenticeships amongst the range of vocational qualifications that they consider. For example, Buscha and Irwin (2013) use administrative data from the Individualised Learner Record (ILR) matched to earnings data from HMRC tax records, to estimate earnings regressions to establish the earnings premium between those who complete and fail to complete an apprenticeship. The regressions control for a range of other factors associated with earnings, including gender, age and ethnicity, regional dummies and for some individuals, prior qualification levels. The results produce very similar estimated earnings differentials to the LFS research discussed above, with those completing a Level 3 apprenticeship earning on average 21% more in the first year after completion and 17% in the fourth year, compared to those who do not achieve. At Level 2, those who complete their apprenticeship also earn around 20% more than non-completers in the first year, falling to 13% after four years.

Bibby *et al* (2014) extend the results of the previous paper by adding more years of data, and in particular by adding more controls for labour market experiences before the apprenticeships were undertaken, in an effort to hold constant more differences between achievers and non-achievers.[4] The results again reveal substantial earnings differentials associated with apprenticeships, even with this expanded set of control variables. Averaged across 3-5 years after the apprenticeship, the earnings differentials are estimated to be 16% and 11% at Level 3 and Level 2 respectively.

A recent CVER discussion paper (Cavaglia *et al.*, 2017) has used the administrative data to focus exclusively on apprenticeships. Their analysis differs from that presented here, in that it focuses on a single cohort of young people, who completed compulsory education in 2003 and who are followed until age 28 in 2015, at which point earnings differentials are estimated. Other key differences are the comparison groups used, and the range of control variables included. Cavaglia *et al.* (2017) compare the earnings of former apprentices at a particular level to individuals without an apprenticeship but qualified to that same level (i.e. Level 2 for the Intermediate Apprenticeship comparison, and Level 3 for the Advanced Apprenticeship comparison). This contrasts to the non-achievers comparison group used here. Their analysis does not consider pre-apprenticeship earnings in a difference-in-differences framework, but does include a wider range of controls, in particular prior attainment from the years of compulsory schooling.[5] Their main results show an annual earnings differential for men with an Intermediate Apprenticeship of 23% at age 28 relative to those men with GCSEs, and 16% relative to those men whose highest qualification is a vocational Level 2 qualification. The equivalent figures for women are 15% and 4% respectively. At Level 3, men with an Advanced Apprenticeship are observed to receive an annual earnings differential at age 28 of 37% relative to men with A levels, and 35% relative to men whose highest qualification is a vocational Level 3 qualification. The equivalent figures for women are 9% and 15% respectively.

Most of the papers discussed above are reports for government departments and policy-makers. Within the more academic literature, there has been focus in some papers on trying to estimate a causal effect of apprenticeships on wages/earnings, that takes account of non-random selection onto apprenticeships, acknowledging that those who undertake and complete an apprenticeship

[4] Specifically, they control for the number of days the individual received benefits in the year before the apprenticeship, whether the individual had an inactive benefit spell in that year, and the number of days in sustained employment the individual had before the apprenticeship.

[5] Such prior attainment information is not used in the main analysis here as it is not available for the sample of older apprentices, though is considered for the younger apprentices only as part of the robustness checks in the results section.

may have higher levels of wage-enhancing characteristics such as ability or motivation. One such example, in an Austrian setting, is Fersterer *et al*. (2008), who make use of the fact that apprentices are employed within firms, and that some firms can go out of business whilst employing apprentices, thus creating exogenous variation in the amount of apprenticeship training received, from the point of view of the apprentice. The authors therefore use time to firm failure as an instrument for the length of apprenticeship completed. Their IV results are slightly higher than, but not significantly different from, the OLS results, suggesting that selection issues into apprenticeships are not strong. The estimated differential is around 4% per year of apprenticeship training. With the typical apprenticeship in Austria lasting around three years, this estimate is similar to the OLS estimates for the UK discussed above.

This paper adds to the literature on apprenticeship returns, to which the papers discussed above belong. We provide the detailed analysis of apprenticeship earnings differentials, with disaggregations by, for example, gender, level, and sector of work, that are expected by policy-makers, whilst also following the academic literature in using econometric techniques to obtain more robust estimates that reduce or eliminate selection effects. Such methods are discussed later, after the description of the data used, which now follows.

3. Data

The analysis makes use of data from the Individualised Learner Record (ILR), which is an administrative data set covering the population of funded learners in Further Education in England. The ILR contains detailed information on the learning undertaken by individuals, including the learning aim, type of qualification, level, subject area, training provider, start and end dates, and attainment markers. Being an administrative data set, the availability of individual-level information is far less extensive, with characteristics of learners largely restricted to gender, age and ethnicity, plus regional area indicators.

For our purposes, a key advantage of the ILR data is that it can be matched to HMRC tax records.[6] Specifically, P14, P45 and P60 forms completed by employers provide accurate information on earnings during the year, and start and end date of periods of employment, for those who change employers during the year. These data were used to create a daily earnings measure, which is preferable to an annual earnings measure since it does not depend on the number of days worked per year, which will vary endogenously across individuals.[7]

The ILR was used to identify all individuals who began an apprenticeship between 2004 and 2013. The start date was determined by the availability of ILR data, while the chosen end date allowed a period of time to observe individuals after their apprenticeship. We keep all of those for whom the start and end dates are observed, who were undertaking a Level 2 (intermediate) or Level 3 (advanced) apprenticeship,[8] and who were in the 19-24 year old or the 25+ age groups. The

[6] The matching was kindly undertaken for us by officials at the Department for Education, with the matched anonymised data set provided to us.

[7] Unfortunately no information on hours of work is included in the tax data, and so we could not derive an hourly wage measure.

[8] It was important to disaggregate by level of apprenticeship when determining value, and there were too few observations for higher level (level 4+) apprenticeships to allow meaningful analysis at such levels.

difference-in-differences methodology used, described in the next section, requires daily earnings to be observed both before and after the apprenticeship, providing a further restriction to the sample.[9]

Figure 1 provides an illustration of the data available to us, using the example of those who completed an apprenticeship in 2008, which is year 0 on the horizontal axis. The black bars show that there are around 46,000 such individuals. The grey bars show the number of these individuals observed in the matched ILR-HMRC data set in each year before and after their apprenticeship. As can be seen, the vast majority are observed in some activity in each year. Specifically, they are observed in employment, in receipt of benefits, in school, in Further Education or an apprenticeship. Those not observed in any of these activities (and so not observed in the data set for that year) are likely to be in employment but earning less than the threshold for paying tax, in Higher Education, or unemployed/inactive but not in receipt of benefits. The light blue bars show the number of the 2008 completers who are observed specifically in employment each year. The figure shows that, post-apprenticeship, most observed individuals are in employment. Slightly lower numbers of the 2008 completers were in employment prior to their apprenticeship, though employment is still the modal activity in each of the four years prior to the apprenticeship. There are also individuals observed on an apprenticeship in the years before 2008 (t=0), to be expected for those apprenticeships with a duration of longer than one year.[10] In addition, smaller numbers of individuals are observed undertaking an apprenticeship after their 2008 completion, who are therefore starting a further apprenticeship, in most cases at a higher level. The darker blue bars indicate the number of observations in each year with observed earnings and employment spell date and so for whom a daily earnings measure could be derived. Comparing the height of the dark blue and light blue bars reveals that there is non-missing earnings data for most person-year observations observed in employment.[11] It is such individuals with earnings data both before and after their apprenticeship who are used in the main manalysis, the methodology of which is discussed in the next section.

4. Methodology

The estimation method used to derive the earnings differentials associated with apprenticeships is a regression-adjusted difference-in-differences (DiD) analysis. This method compares the change in earnings for a treated group of apprentices to the change in earning for a control group.

The treatment group were all those whose learning aim was a Level 2 or Level 3 apprenticeship in the period of observation, and who completed that apprenticeship. When choosing a control group, it must be borne in mind the nature of the data set, which was administrative records of all those in Further Education, matched to tax records. The implication of this is that all individuals in the sample are Further Education learners, and hence there are no non-leaners with whom to create a control group. To find a control group to compare to the treated apprentices, we therefore consider those

[9] Hence the reason we did not consider the 16-18 year old apprentices was that they were not observed in employment prior to their apprenticeship.

[10] The mean duration of observed apprenticeships is around one year, though even for these, this one year could be observed in two consecutive financial/tax years. The median number of tax years in which an apprenticeship is observed in the data set is two.

[11] Note the activity bars in any one year can sum to more than the total number of observations for that year, since individuals can be in multiple activities, for example in employment and also in receipt of in-work benefits.

who began an apprenticeship but did not complete. Using such a comparison group has the advantage that both treatment and control groups have chosen to do, and have been accepted onto, an apprenticeship. This therefore avoids selection effects onto apprenticeships, which could bias results if the factors that determine selection into and acceptance onto an apprenticeship are correlated with earnings. The clear disadvantage of the method, however, is that failing to complete an apprenticeship is not random. While some reasons for terminating an apprenticeship may be exogenous (for example, a sponsoring firm going out of business, a college course being not what was expected, unexpected differences in teacher/instructor quality etc), there will be individual characteristics, such as motivation and ability, that may also affect the likelihood of completing an apprenticeship, and which are also determinants of earnings, and so have the potential to bias results.[12]

The use of the DiD framework is therefore important to mitigate such biases as far as possible. This method compares the change in earnings before and after an apprenticeship, for the treatment group and the control group. By controlling for pre-apprenticeship earnings, the method ensures that any differences in unobserved characteristics between completers and non-completers that influence earnings are differenced out, to the extent that such characteristics remain constant over time. The method also controls for anything else that changes over time that might affect earnings, to the extent that such changes affect treatment and control groups equally. Though it is possible that other biases that do not fulfil these conditions may remain, it can also be argued that since the primary results of interest are a comparison between the estimated effects for 25+ and 19-24 year old apprentices, then as long as any remaining biases affect each group equally, they will be differenced out when making the comparison between the two groups. In effect, the analysis is actually using a triple-difference methodology.

Setting the DiD analysis within a regression framework also allows us to control directly for observable characteristics such as duration of apprenticeship, age completed the apprenticeship, ethnicity, and current tax year dummies for observed earnings. The estimated equation therefore takes the form:

$$\log(earnings)_{it} = \alpha + \beta T_i + \gamma POST_t + \tau(T * Post)_{it} + \theta X_{it} + u_{it}$$

The time indicator, t, indicates the number of tax years prior to (if t is negative) the tax year in which the apprenticeship began or since (if t is positive) the tax year in which the apprenticeship was completed. Observations where t=0 are excluded, which will often be multiple tax years, as discussed in the previous section, since even a one year apprenticeship will often extend over two tax years.[13] The variable POST indicates that t>0. T is the treatment indicator for those who completed their apprenticeship. The DiD coefficient is therefore τ, the coefficient on the interaction term between the treatment group and treatment-on (POST) indicators. The X vector contains the control variables listed above.

Considering the length of the window in which we observe individuals in employment, there is a trade-off when using wider windows. On the one hand, observing individuals for longer periods before and after their apprenticeship will provide more information on slightly longer time impacts of apprenticeships while also averaging out short-term transitory variation. On the other hand, using

[12] Unfortunately, we do not observe the reason for non-completion.

[13] Tax years spent even partially on the apprenticeship itself are not used in the analysis

wider windows makes greater demands on the data, requiring individuals to be observed in paid employment for more successive years, and therefore reduces the sample size. We also have to take into account that one of our age groups are relatively young and so will have fewer years of pre-apprenticeship employment. The widest window we consider is therefore three years either side of the apprenticeship, while we also check the robustness of the results to using only a 2 year or 1 year window either side. The analysis is undertaken separately by gender and by level of apprenticeship. There are therefore 24 estimated earnings differentials produced (2 age groups x 2 genders x 2 apprenticeship levels x 3 time windows).

In an attempt to increase the number of observations, we also tried relaxing the requirement that individuals have to be observed in *each* year within the given window either side of the apprenticeship, and re-estimated the results using any individual who undertook an apprenticeship and for whom employment wages are observed at least once, to check the robustness of the results to doing this.

5. Results

5.1 Main Results

Before looking at the results of the empirical analysis, Figures 2 and 3 show, for the 19-24 year old and 25+ age groups respectively, the average log real daily earnings amongst individuals who undertook an apprenticeship, in each year before and after their apprenticeship. Completers and non-completers are shown on the same diagram by solid and dashed lines respectively, while apprenticeships at Intermediate and Advanced Levels are shown on separate diagrams within the same figure.

Figure 2 shows, for the younger age group taking Intermediate Apprenticeships, that for both genders, real earnings were already rising over time before the apprenticeship was undertaken, as would be expected for a young age group in their formative years in the labour market. For both genders, those who went on to complete their apprenticeship were already earning more pre-apprenticeship than those who would start but not complete an apprenticeship. This suggests ability differences between the two groups, and shows the importance of controlling for pre-apprenticeship earnings. The key point, however, is that the gap between the earnings of the treatment and control groups widens over the period that the apprenticeship is undertaken, showing the added value of completing an Intermediate Apprenticeship. There is a similar pattern in the lower diagram in Figure 2 for Advanced Apprenticeships taken by the younger age group, with the difference that pre-apprenticeship earnings are the same for completers and non-completers. In this case the earnings gap only opens once the apprenticeship is completed.

Figure 3 displays the same diagrams for the older age group, and reveals significant differences. For the age 25+ apprentices, real earnings were not rising over time prior to the apprenticeship, but were fairly constant. At the Intermediate level, the earnings profiles over time also remain fairly flat after the apprenticeship, so that there is little evidence for a boost in earnings, either in absolute terms or relative the control group of non-completers. At the Advanced level there is some evidence of an uplift in earnings after successfully completing an apprenticeship, so that the earnings gap between treatment and control groups widens. As an aside, it is also of interest to note the much

larger gender gap in wages observed for this group, compared to that experienced by the 19-24 year old apprentices.

Table 1 displays the results from the empirical analysis based on the restricted sample where each individual needs to be observed in paid employment in every year in the relevant window, in order to be included. The table reports only the DiD coefficients, τ, with each estimate in the table coming from a different regression.[14]

The first thing to note is that every coefficient is positive and statistically significant. Thus, in every case (gender, age group, apprenticeship level and sample window considered), a completed apprenticeship is associated with a larger increase in earnings, relative to their pre-apprenticeship level, than a non-completed apprenticeship. This demonstrates the continued value placed on apprenticeships by the labour market. The estimated earnings differentials are mostly smaller than those typically observed in the literature, as summarised in Section 2 above, though most of those estimates were based on simple cross-sectional equations, and did not use information on prior earnings to difference out the effects of unobserved characteristics that remain constant over time.

The main aim of the analysis is to compare the gain in earnings for older (25+) apprentices and those aged 19-24. In every one of the twelve comparisons in Table 1, the estimated earnings differential is higher for the 19-24 year olds group. For 10 of the 12 comparisons, this difference is statistically significant (the two exceptions being for Intermediate Apprenticeships for both men and women, when the two year window is used either side of the apprenticeship). In most cases, the apprenticeship earnings differential for 19-24 year olds is around twice the size of the differential for those aged 25+, peaking at almost four times as large in the case of Advanced Apprenticeships for men using the 2-year window. There is therefore strong and robust evidence that earnings rise more after an apprenticeship when undertaken between the ages of 19-24, compared to when undertaken at the age of 25 or later.

Briefly considering the other finding revealed by Table 1, earnings differentials are usually, though not always, larger for Advanced compared to Intermediate Apprenticeships, and for men compared to women. The higher male differentials are, on the whole, observed for both levels of apprenticeship and for both age groups. The choice of length of window does not have a consistent effect on the point estimates of the differentials across categories, but the pattern of results is robust to differences in window length.

Table 2 shows the results when we relax the sample requirement that an individual must be observed with earnings information in every year, before and after the apprenticeship, of the chosen window. Any individual with at least one observation on earnings is included. This has the benefit of greatly increasing the number of observations used for analysis, at the cost of introducing more noise into the estimates since it is not necessarily the same individuals who are observed before and after the apprenticeships. The aim of this analysis is not to put too much attention on the results, but rather to test the robustness of the main results in Table 1 to this change in sample. If the results can be shown to be robust then this is of benefit since it supports the arguments that (i) the main results are not specific only to the sub-sample of individuals with continuous employment, but can be generalised, and (ii) when undertaking the disaggregated analysis by Apprenticeship Framework later that we can use the larger, unrestricted sample without fear of affecting the results.

[14] Full results showing coefficients on all variables are available from the authors on request.

The results in Table 2 show that the results are indeed qualitatively robust to using the unrestricted sample. In every one of the twelve pairs of estimates, it is again the case that the estimated wage differential is higher for the 19-24 year olds completing an apprenticeship than for those ages 25+. In most cases, the differential for the younger group is at least twice as large, which together with the lower standard errors that come with the larger sample size, means that every difference in differentials between age groups is statistically significant. It is also the case that differentials remain higher for males than the equivalent differentials for females in most cases, the exceptions being Advanced Apprenticeships amongst the older age group in each estimated window. In every case, the differential associated with an Advanced Apprenticeship is larger than the differential for an Intermediate Apprenticeship for the same group. We can therefore have confidence when we use the unrestricted sample later for the disaggregated results.

5.2 Further Robustness Checks

In this section, we consider alternative specifications for estimating the earnings differential, to determine whether the main results discussed above are due to the particular specification estimated, or are robust to specification changes.

The first issue to consider is that the 25+ age group clearly contains a wide range of ages. It is more likely that older members of this age group will differ to 19-24 year old individuals in terms of their unobserved characteristics, coming as they do from older cohorts who entered the labour market in different eras, and therefore potentially having different attitudes, motivation etc. The younger members of the age 25+ group are likely to be more similar to the 19-24 years olds in terms of their unobserved characteristics. We therefore re-estimated the DiD analysis undertaken above, this time on two equal-sized aged groups either side of the age 25 cut-off, comparing 20-24 year olds to 25-29 year olds. The unrestricted sample was used for this analysis, to ensure large enough sample sizes for the 25-29 year old group, appealing to the result shown in the previous sub-section that the unrestricted sub-sample gives qualitatively similar results to the restricted sample. The results are reported in Table 3, and show that it remains the case that in every one of the twelve pairs of estimates (disaggregated again by level, gender, and length of observation window) the estimated earnings differential associated with apprenticeship completion is larger for the 19-14 year old group than for the age 25+ group. It is true that for the Intermediate Apprenticeships, the differences in the differentials between age groups are smaller than those observed above using the full sample, and none of these differences are statistically significant. For the Advanced Apprenticeships for males, however, the differential for 20-24 year olds remains around twice that for 25-29 year olds, with all differences being statistically significant, while for women the differences are smaller, but are again statistically significant in the case of the 3-year and 2-year observation windows.[15]

A key robustness check to run when conducting a difference-in-differences analysis is to conduct a placebo treatment test, whereby the before-after analysis is estimated around a date before the

[15] It is this consistent pattern of results that is of interest here, rather than the actual values. The main results in the previous section therefore remain the ones of interest, estimated as they are on the older individuals actually doing the apprenticeships, as opposed to those falling only into the 25-29 age range. The robustness check here shows that these earlier results were not being driven by something fundamentally different between older and younger workers, since the same pattern is observed when comparing adjacent age groups on either side of the age 25 cut-off.

individuals actually undertook their apprenticeship. The idea is that a DiD estimate around this placebo treatment date should yield statistically insignificant coefficients, which would show that the growth of earnings over time were on the same common trend (i.e. changing by the same amount) for the treatment and control groups, before any apprenticeships were undertaken. In particular, the placebo treatment was assumed to occur between t=-1 and t=-2, so that a full year of earnings are observed before the individual actually undertook their apprenticeship. The results are shown in Table 3, using 3-year and 2-year windows and the main sample, that is the sample restricted to individuals with observed earnings in all years of the window, either side of the 'fake' apprenticeship.

The results show that in the case of males, the DiD estimates around these 'fake' apprenticeships are all statistically insignificant, with very small point estimates (8 estimates: 2 age groups x 2 apprenticeship levels x 2 window lengths). There is therefore no evidence for differential changes in earnings between achievers and non-achievers before the apprenticeship was started, in the case of males. For female apprentices, all four estimates of the effect of the placebo treatment for the older age group are similarly statistically insignificant. For the younger females, there is some evidence of faster earnings growth for the treatment group compared to the control group even before the apprenticeships were started, with estimates for both apprenticeship levels being statistically significant when using the three year windows around the fake treatment, though note that when the two-year windows around the fake treatment are used, all estimates for young females become statistically insignificant. For females, we therefore cannot absolutely rule out (based on the three-year window results) the possibility that the higher earnings differential for the younger age group relative to the older age group would have occurred anyway, because of this prior earnings growth for the younger female achievers. Even if we were to subtract these significant differentials from the placebo tests for young female apprentices (from Table 4), as an estimate of the faster earnings growth that would have occurred for female young achievers anyway, away from the actual differentials observed for such young female apprentices (in Table 1), they are still higher than the observed differentials for the older female age group in Table 1, in each case.

A third robustness test involved using an alternative methodology, specifically estimating a treatment effect using a propensity score matching methodology rather than a DiD analysis. The treatment and control groups remain the same as above, namely those who complete and fail to complete, respectively, their apprenticeships. The matching is conducted using the following characteristics, to ensure that the comparison is between individuals who are as similar as possible in terms of their observed characteristics: apprenticeship duration, the year the apprenticeship was completed, the age at which the apprenticeship was completed, and ethnicity. In addition, and crucially, we also match on labour market outcomes prior to the apprenticeship, namely total real annual pay, total days employed, and total days in receipt of benefits in the year prior to beginning the apprenticeship. The treatment effects are estimated using the five nearest neighbours and a 0.005 caliper, with the outcome variable being the log of daily earnings observed in the first full tax year after the apprenticeship has been completed. Standard errors are bootstrapped to allow for the fact that the propensity score is estimated rather than known. The analysis is therefore comparing individuals with the same observed characteristics and the same history of labour market experiences before they began an

apprenticeship, however one group complete that apprenticeship and the other group do not.

The results are reported in Table 5, and show exactly the same patterns as the DiD results in Tables 1 and 2, discussed above. In every pair of cases in Table 5, the estimated treatment effect (estimated earnings differential) is larger for the 19-24 year old group than for the age 25+ group, usually around twice as large, and up to three times as large in one case. It therefore remains the case that, when matching on observed characteristics and prior labour market history, those who complete their apprenticeship subsequently earn a higher wage than those who do not, with this gain in value being larger for younger than for older apprentices. The figures in the lower rows of Table 5 show that the matching exercise was successful, with the median/mean % standardised bias[16] being much reduced in the matched sample relative to the unmatched sample. Similarly, Rubin's B statistic[17] is far below the value of 25 in every case, while Rubin's R statistic[18] is close to ideal value of 1 in every case, and easily within the bounds of 0.5-2.0 expected of a good match. The matched sample of treated and control group observations therefore seem to be well balanced in each of the cases considered.

A final robustness check involved determining the robustness of the results to the choice of control group. As explained earlier, the choice of the 'non-achievers' control group was due to the nature of the administrative data set available, in which all respondents are by definition learners, given that the ILR forms the basis of the data set. In addition, the ILR does not contain information on prior attainment, so that we cannot compare learners to those one level lower in the qualifications hierarchy. Due to the potential for systematic differences in unobserved characteristics between those who successfully complete and those who do not complete an apprenticeship, then such differences could also contribute to any observed differences in earnings. As argued earlier, the effects of such bias are mitigated by the DiD methodology employed, so that any differences in characteristics between groups that remain fixed over time and which affect earnings equally over time, will be differenced out and so not influence the estimated effect of the apprenticeships. In addition, any biases that remain after the DiD analysis within age groups will be differenced out of the comparison between age groups, to the extent that they affect both age groups equally. Nevertheless, it would be of interest to determine whether the use of the non-achievers control group is particularly prone to producing biased results, compared to the 'level below' control group more commonly found in the literature.

For the young age group only, it is possible to match the ILR data to data found in the National Pupil Database (NPD), which contains information on all national tests at ages 11,

[16] The % standardised bias is the percentage difference between the sample means of an explanatory variable in the treated and non-treated samples, with the mean/median % standardised bias being the average bias across all explanatory variables. These and the following statistics were obtained using the *pstest* procedure, developed in Stata by Edwin Leuven and Barbara Sianesi.

[17] Rubin's B statistic is the absolute standardised difference of the means of the linear index of the propensity score in the matched treated and non-treated groups.

[18] Rubin's R statistic is the ratio of the variances of the propensity score index in the matched treated and non-treated groups.

14 and 16 taken by the population of school pupils.[19] This allows us to do two things with the younger age group. First, through the NPD and ILR combined, and knowledge of who proceeded to Higher Education,[20] we know the highest level of attainment for all individuals. We can therefore conduct a 'traditional' analysis of earnings differentials, comparing the earnings of those for whom the apprenticeship is their highest qualification, to those whose highest level of attainment is one level below the apprenticeship being considered. Second, we can add controls for school-level test scores, as a proxy for ability, which may pick up some of any remaining differences in unobserved characteristics between treatment and control groups.

We therefore estimate cross-sectional earnings equations, for the sample who undertook their apprenticeship between the ages of 19-24. The dependent variable is log daily wages in 2015. For each apprenticeship level/gender combination, two specifications are estimated (thus eight specifications overall), one with the non-achievers who registered for the same apprenticeship as the comparison group, and the other where the comparison group comprises those individuals whose highest qualification is one level below the apprenticeship of interest. The estimated equations control for ethnicity, cohort fixed effects, and time elapsed since the highest qualification was completed. The results are reported in Table 6. The estimated coefficients are larger than the DiD effects estimated above, as was expected since these are simply cross-sectional relationships with no control for pre-apprenticeship labour market outcomes.[21] The important point is the comparison in results between control groups. In all four pairs of estimates, the wage differential estimated using the 'level below' comparison group is much larger than the one estimated using the non-achievers control group. If we then add school fixed effects, local education authority fixed effects, and controls for school level attainment from the NPD[22], to proxy unobserved ability of the individuals and compensate for the absence of the pre-apprenticeship labour market outcomes, the results (see Table 7) are lower again, and back in the range estimated using the DiD framework. Crucially, the estimated earnings differentials are similar across the two control groups, in each of the four pairs of equations, with the 'level below' control group producing the slightly higher estimate in three of the four cases. Thus there does not seem to be anything in the use of the non-achievers control group that is artificially inflating the estimated earnings differentials, and we can conclude that the main results discussed above are not an outcome of using this control group, rather than the more traditional approach comparing apprentices to those qualified one level below.[23]

[19] NPD information is not available for many older individuals in the 25+ age group, since the NPD only supplies data from 2002 onwards. It was therefore not possible to use this data source for the main analysis.

[20] The data set used is the LEO (Longitudinal Education Outcomes) data set, which combines data from the NPD, ILR and HESA (Higher Education Statistics Agency), with HMRC tax records in the WBL data set used in the main analysis.

[21] The 'level below' control group does not lend itself to a DiD framework, so for the purposes of comparison, both equations are estimated by OLS.

[22] Specifically, the equations additionally control for GCSE performance at age 16 in Maths and English as well as overall number of GCSEs obtained, performance in Key Stage 3 Maths and English tests at age 14, performance in Key Stage 2 Maths and English tests at age 11, and whether the individual was eligible for free school meals, as an indicator of family background.

[23] It is also worth noting here, that the 'level below' specifications in Table 7 are much more similar (rather than the DiD estimates above) to the specifications estimated by Cavaglia *et al.* (2017) in an earlier CVER discussion paper, in that they do not use non-achievers in the comparison group, and they do use prior

5.3 Disaggregated Analysis by Apprenticeship Framework

The results in the previous two sub-sections have established first that the earnings differential received by older apprentices, relative to apprenticeship non-completers, is smaller than that received by younger apprentices, this being the case for every combination of apprentice gender, level of apprenticeship and duration of observation window considered, and second that this result is robust to choice of age group, control group and methodology used. An important question for apprentices, firms and policy-makers alike, however, is to what extent this result is due to older apprentices earning a smaller differential than younger apprentices within the same Apprenticeship Framework, and to what extent it is due to older apprentices typically undertaking apprenticeships in frameworks that earn lower differentials, on average. To answer this question, we therefore undertake a disaggregated analysis by Apprenticeship Framework in which an apprenticeship was taken. In order to maximise sample sizes in this disaggregated analysis, the unrestricted sample from the earlier section was used, in which there was no requirement for an individual to be observed in all of the years of the observation window.

The choice of framework pursued by apprentices is dominated by a few with very large numbers, such as Business Administration, Customer Services, and Health and Social Care. While such frameworks had sufficient observations with which to estimate equations on their own, many others were too small in terms of number of observations. We therefore grouped frameworks together into groups that were broadly consistent with Sector Skills Councils. This left us with 20 categories, including an 'other' category which contained all those frameworks with too few observations even when grouped together at Sector Skills Council level.[24]

To estimate the earnings differentials by framework, the following extension to the DiD model was estimated:

$$\log(earnings)_{it} = \alpha + \beta_1 T_i + \beta_2 POST_t + \beta_3 (T * Post)_{it} + \sum_{k=1}^{19} \delta^{(k)} F_i + \sum_{k=1}^{19} \gamma_1^{(k)} \left(F_i^{(k)} * T_i\right) + \sum_{k=1}^{19} \gamma_2^{(k)} \left(F_i^{(k)} * POST_t\right) + \sum_{k=1}^{19} \gamma_3^{(k)} \left(F_i^{(k)} * T_i * POST_t\right) + \theta X_{it} + u_{it}$$

attainment controls. Even though some differences remain between these specifications and those of Cavaglia *et al.* (2017), in terms of cohorts studied and comparison groups being one level below rather than at the same level, the findings are quite similar. In particular, Cavaglia *et al.*'s (2017) findings of much larger differentials for men than for woman, particularly for Advanced Apprenticeships, are replicated in Table 7.

[24] The 20 categories are, in order of number of observations: Business Administration, Custome Services, Health and Social Care, Hospitality and Catering, Retailing, Hair and Beauty, Child Development, Construction, Engineering, Other, Automotive, IT, Sport and Active Leisure, Electrotechnical, Logistics, Land-Based, Plumbing, Manufacturing, Travel Leisure and Tourism, and Accountancy.

$F_i^{(k)}$ is an indicator equal to 1 if individual i's apprenticeship was in framework k, for k=1,…, 20 frameworks. The base group is the 'Other' category. These dummies are then fully interacted with the variables indicating treatment and post-treatment. For any apprenticeship framework k, the DiD estimate of the earning differential associated with completing an apprenticeship in that framework, relative to a non-achiever in the same framework, is given by $\beta_3 + \gamma_3^{(k)}$. The equation was estimated separately for each combination of gender and level, using the three-year observation window. The results, for the DiD estimates, are displayed in Figures 4 and 5 for males and females respectively, with separate diagrams for Intermediate and Advanced Apprenticeships within each figure.

Since cell sizes within frameworks are quite small in a number of cases, the error bars showing 95% confidence intervals around the DiD estimates are quite wide in those cases. It is therefore hard to determine statistically significant differences in the estimated earnings differentials, either between frameworks, or between age groups within frameworks. A broad summarising of the evidence would say that there appear to be many more frameworks where the estimated differential for the 19-24 year old age group is larger than that for the age 25+ group. In some cases, these differences between age groups within frameworks are statistically significant, such as Business Administration and Accountancy at the Intermediate level for males, plus Manufacturing and Travel and Tourism at both levels, and similarly Business Administration at both levels for females, plus IT at the Intermediate level and Engineering at the Advanced level. If there is one diagram where there appear to be fewer frameworks with higher earnings differentials for younger apprentices, it would probably be the male Advanced Apprenticeship diagram.

To undertake a more formal analysis of these differences, we performed a sectoral decomposition analysis across the 20 framework categories, to determine in each case the proportion of the young-old difference in earnings differentials that occurs within frameworks and the proportion that occurs between frameworks, as shown in the equation below:

$$\Delta D = \sum_k \Delta D_k \overline{S_k} + \sum_k \Delta S_k \overline{D_k}$$

where Δ indicates a difference between old and young age groups, D is an earnings differential, and S is proportion of apprentices in framework k, with a bar indicating an average value across age groups. The first term on the right hand side represents the within framework differences in earnings differentials between older and younger age groups, while the second term is the between framework differences, showing the differences between groups in the number following each framework weighted by the average earnings differential in that framework. Note that the left-hand side variable of this equation is the weighted average of the difference in earnings differentials between groups across frameworks, and so not the same as the DiD estimate of the difference in earnings differentials between groups estimated using the pooled sample above.

Table 8 reports the proportion of the variation in earnings differentials between age groups that occurs within and between frameworks, for each combination of gender and apprenticeship level. The table makes clear that, for three of the four cases, that most of the variation is occurring *within* frameworks. Thus, in most cases, much of the reason for the lower earnings differential received by older apprentices upon completion of their apprenticeship is due to them receiving a lower differential than younger apprentices in the same subject area. The exception, and as suggested by Figures 4 and 5 above, is the Advanced Apprenticeships for males, where two-thirds of the variation

has occurred *between* frameworks. In this case, therefore, it seems that older male apprentices are choosing to undertake Advanced Apprenticeships in sectors that have lower earnings differentials on average, such as Business Administration.

6. Conclusion

There has been a huge growth in the number of people undertaking an apprenticeship in England over the last decade or so. With this increase in supply of successful apprentices onto the labour market, we might expect to see variation in the earnings differentials that they receive, perhaps not matching those achieved by earlier cohorts of apprentices, who had more scarcity value in the labour market.

Our results suggest that this is not the case, though. We use difference-in-differences techniques, that compare the change in daily earnings before and after an apprenticeship is undertaken, for a treatment group of individuals who complete their apprenticeship relative to a control group who do not complete an apprenticeship at the same level. Such methods control for unobserved differences between treatment and control groups that affect labour market outcomes (to the extent that they are captured by prior experiences before the apprenticeship is undertaken), as well as controlling for other factors that affect the change in earnings over time (to the extent that they affect both treatment and control groups equally). The outcome is that completing an apprenticeship is still associated with a significant increase in earnings amongst recent apprentices.

There is variation in the size of this differential, however. The earnings differential is, in most cases, larger for a Level 3 (Advanced) Apprenticeship, than for a Level 2 (Intermediate) Apprenticeship for the same gender and age group. The differential is also consistently larger for males than for females in each comparison pairing, though only significantly so in the case of Level 3 Apprenticeships for 19-24 year olds. The variation in earnings differentials focussed on in this paper, however, was that between the older and younger age groups, with the younger apprentices consistently receiving a higher differential, often around twice as large. This has important implications, given that the fastest growth in apprentice numbers has been amongst the 25+ age group. It is therefore important that further research is undertaken, to explain and understand this differential.

The current research makes a first attempt at explaining the variation, evaluating one possible factor. In particular, we examine whether older apprentices receive a lower earnings differential, on average, because they trained in areas where the differential is lower for everyone, rather than receiving a lower differential than younger apprentices for the same type of apprenticeship (i.e. whether the variation is between or within frameworks respectively). The results show that for women at both apprenticeship levels, and men at Level 2, it is the latter effect that is dominating, and the older apprentices are receiving lower differentials than younger apprentices within the same framework. It is important to determine why the training that they are receiving is apparently less valued in the labour market. This is likely to involve case studies of the training being undertaken, in terms of quality and duration, certification etc. One possibility is that older apprentices are more likely to already work for their training employer before starting their

apprenticeship. The training that they receive is therefore more likely to be top-up training, rather than training to develop new skills for a new career.

For male apprentices at Level 3, it is the case that they are undertaking apprenticeships in frameworks that witness lower earnings differentials on average, such as Business Administration. One issue is therefore whether they are aware of higher differentials available in frameworks such as Construction and Manufacturing. Of course, it could be the case that office workers taking Business Administration Apprenticeships are aware of such differentials, but do not see themselves as working in such sectors, and rather see their apprenticeship as a means to furthering a non-manual office-based career. From an individual point of view, there is no reason why they should not do so, of course, as long as it was a well-informed decision. But from a national point of view, it is important that the growth in apprentice numbers does not continue to be dominated by older apprentices in such lower value frameworks, if apprenticeships are to provide the high-value added training required to raise technical skill levels in the economy.

References

Bibby, D., Buscha, F., Cerqua, A., Thomson, D. and Urwin, P. (2014). 'Estimation of the Labour Market Returns to Qualifications Gained in English Further Education.' Department for Business, Innovation and Skills Research Paper 195.

BIS (2011). 'Measuring the Economic Impact of Further Education.' Department for Business, Innovation and Skills Research Paper 38.

Buscha, F. and Urwin, P. (2013). 'Estimating the Labour Market Returns to Qualifications Gained in English Further Education using the Individualised Learner Record (ILR).' London Department for Business, Innovation and Skills. Department for Business, Innovation and Skills Research Paper.

Cavaglia, C., McNally, S. and Ventura, G. (2017). 'Apprenticeships for Young People in England: Is There a Payoff?' CVER Discussion Paper 010.

Conlon, G., Patrignani, P. and Chapman, J. (2011). 'Returns to Intermediate and Low Level Vocational Qualifications.' Department for Business, Innovation and Skills Research Paper 53.

Fersterer, J., Pischke, J-S. and Winter-Ebmer, R. (2008). 'Returns to Apprenticeship Training in Austria: Evidence from Failed Firms.' *The Scandinavian Journal of Economics,* 110, 733-753.

Hogarth, T. and Hasluck, C. (2003). 'Net Costs of Modern Apprenticeship Training to Employers.' DfES Research Report 418.

McIntosh, S. (2007). 'A Cost-Benefit Analysis of Apprenticeships and Other Vocational Qualifications.' DfES Research Report 834.

National Audit Office (2012). *Estimating Economic Benefits from Apprenticeships*. London: National Audit Office.

Figure 1: Activity of Apprentices in the years around the Apprenticeship

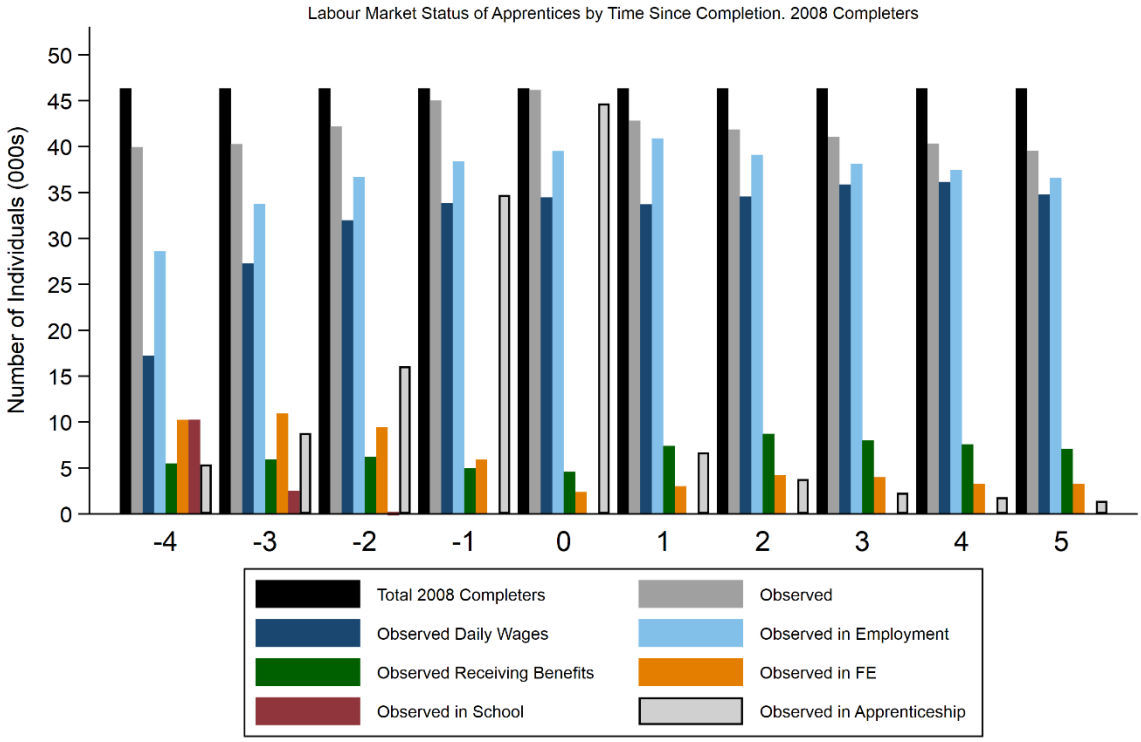


Figure 2: Log Daily Wages of Apprentices who Started Aged 19-24

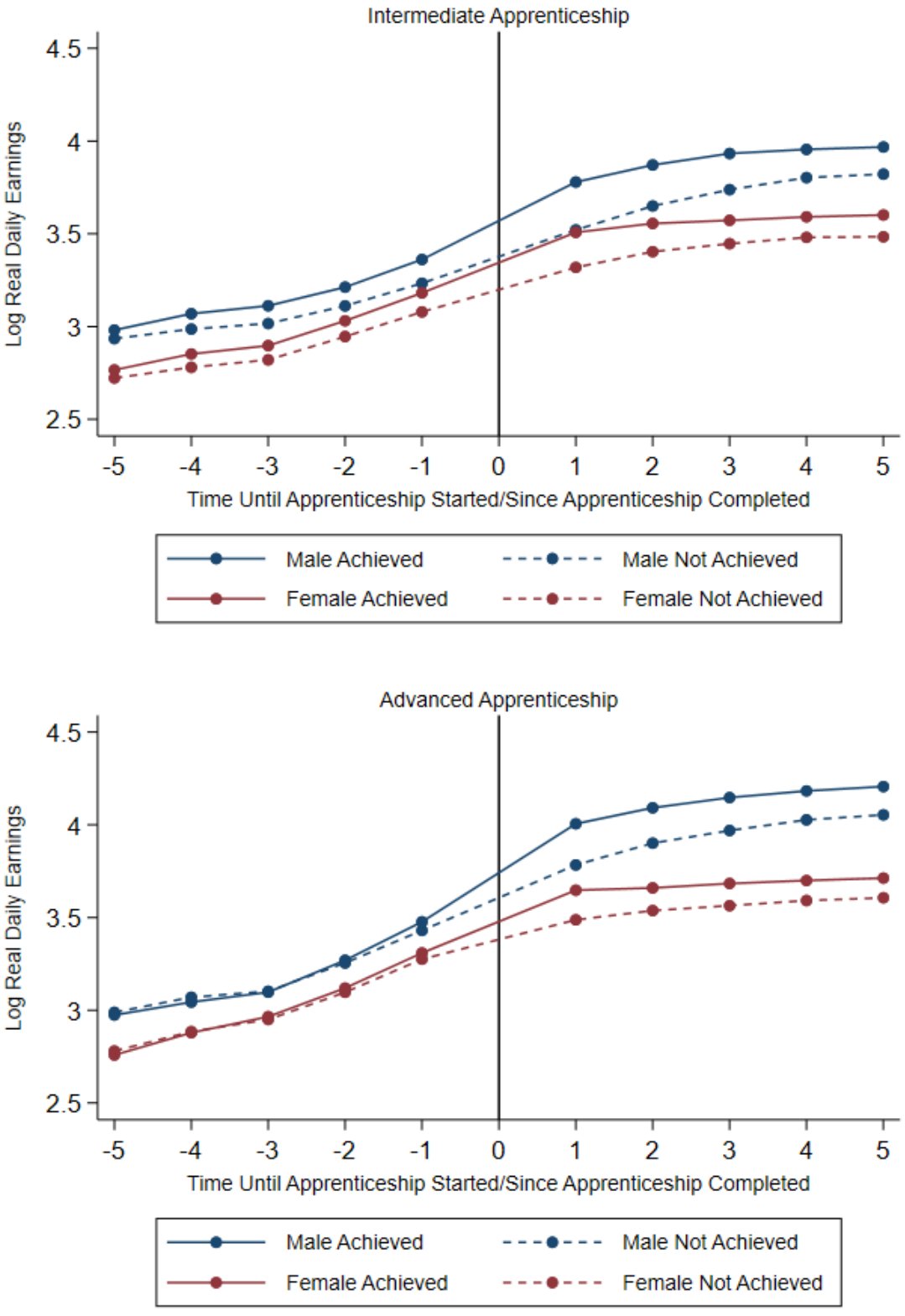


Figure 3: Log Daily Wages of Apprentices who Started Aged 25+

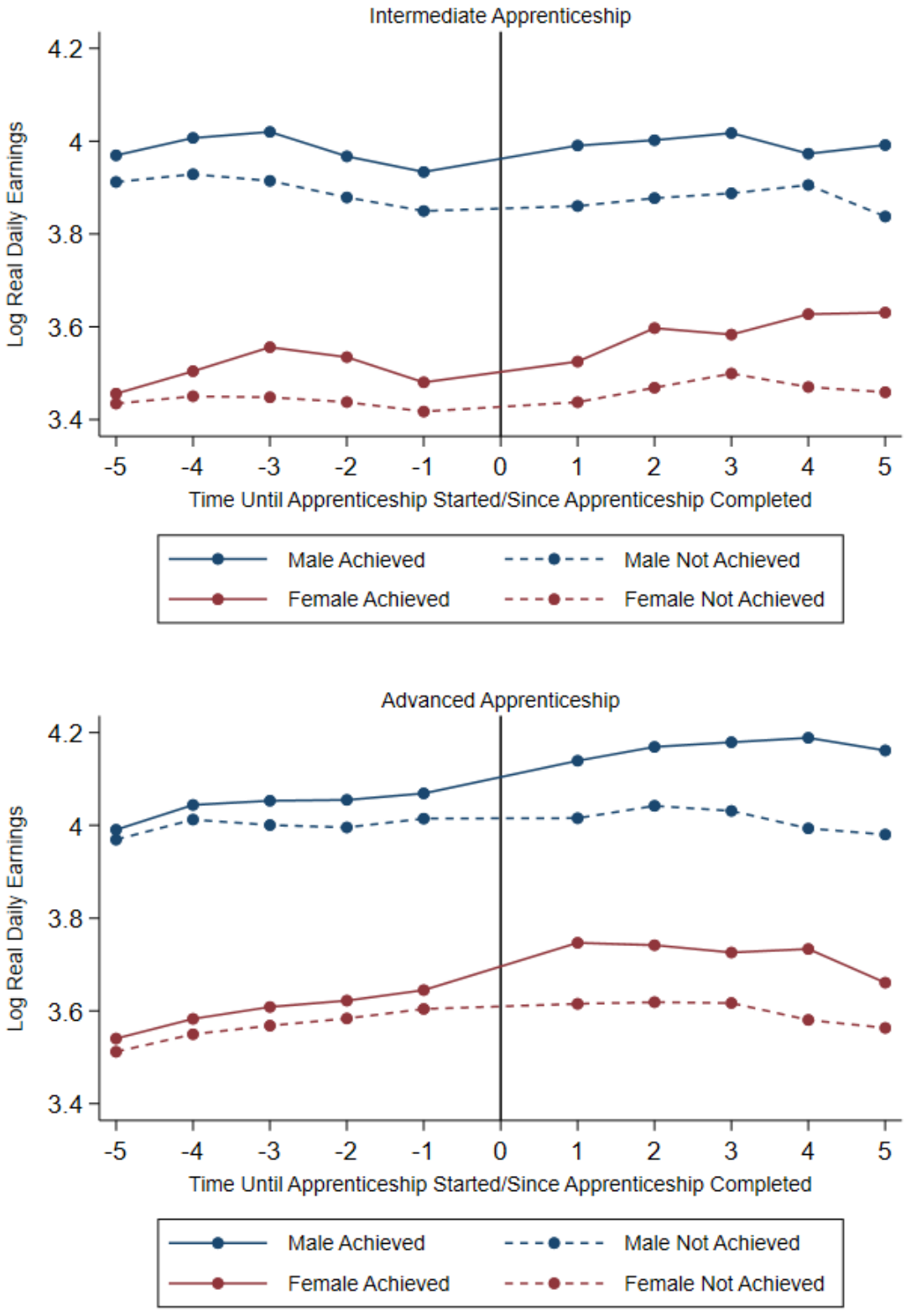


Figure 4: Apprenticeship Returns by Framework - Males

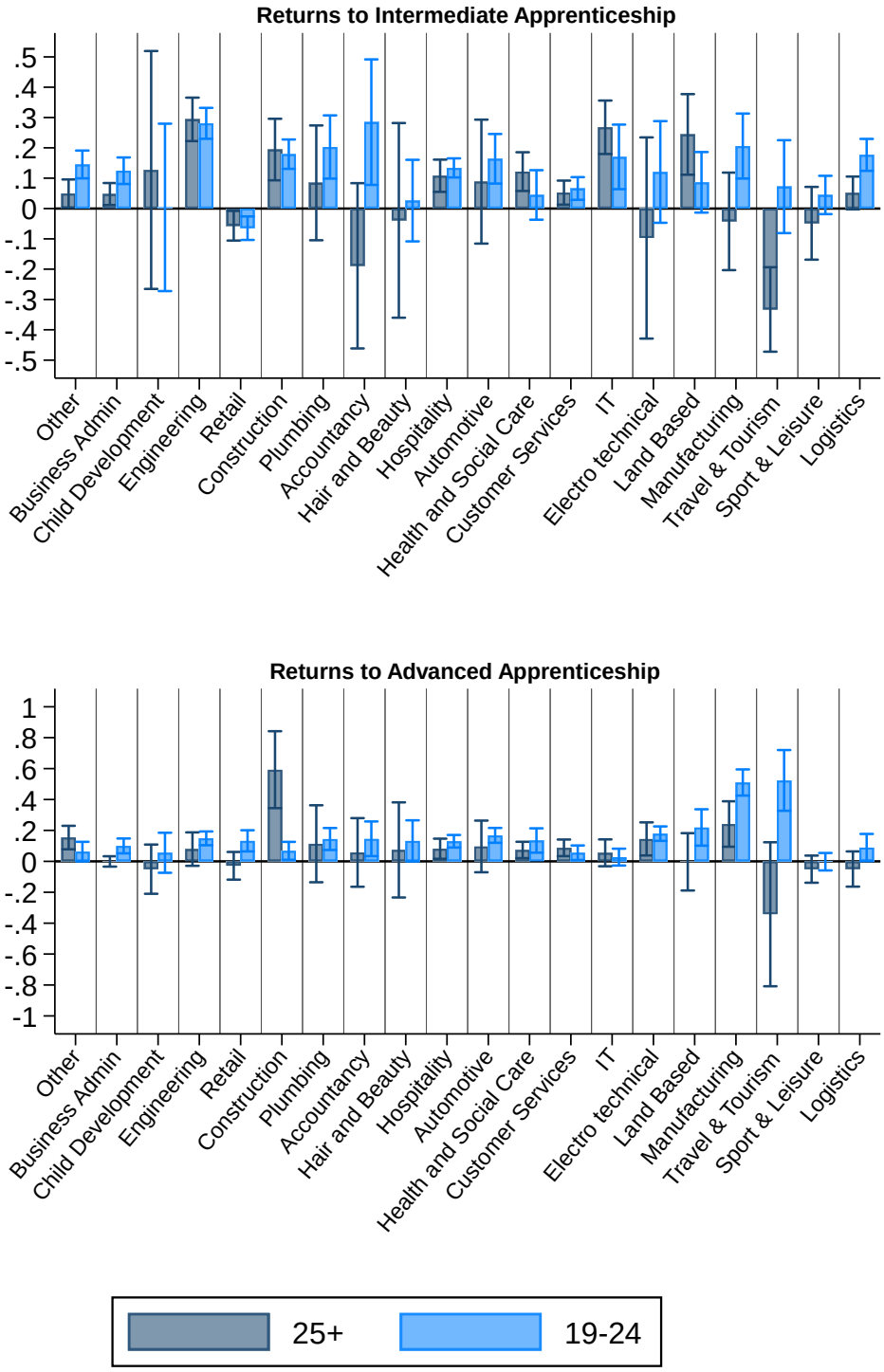
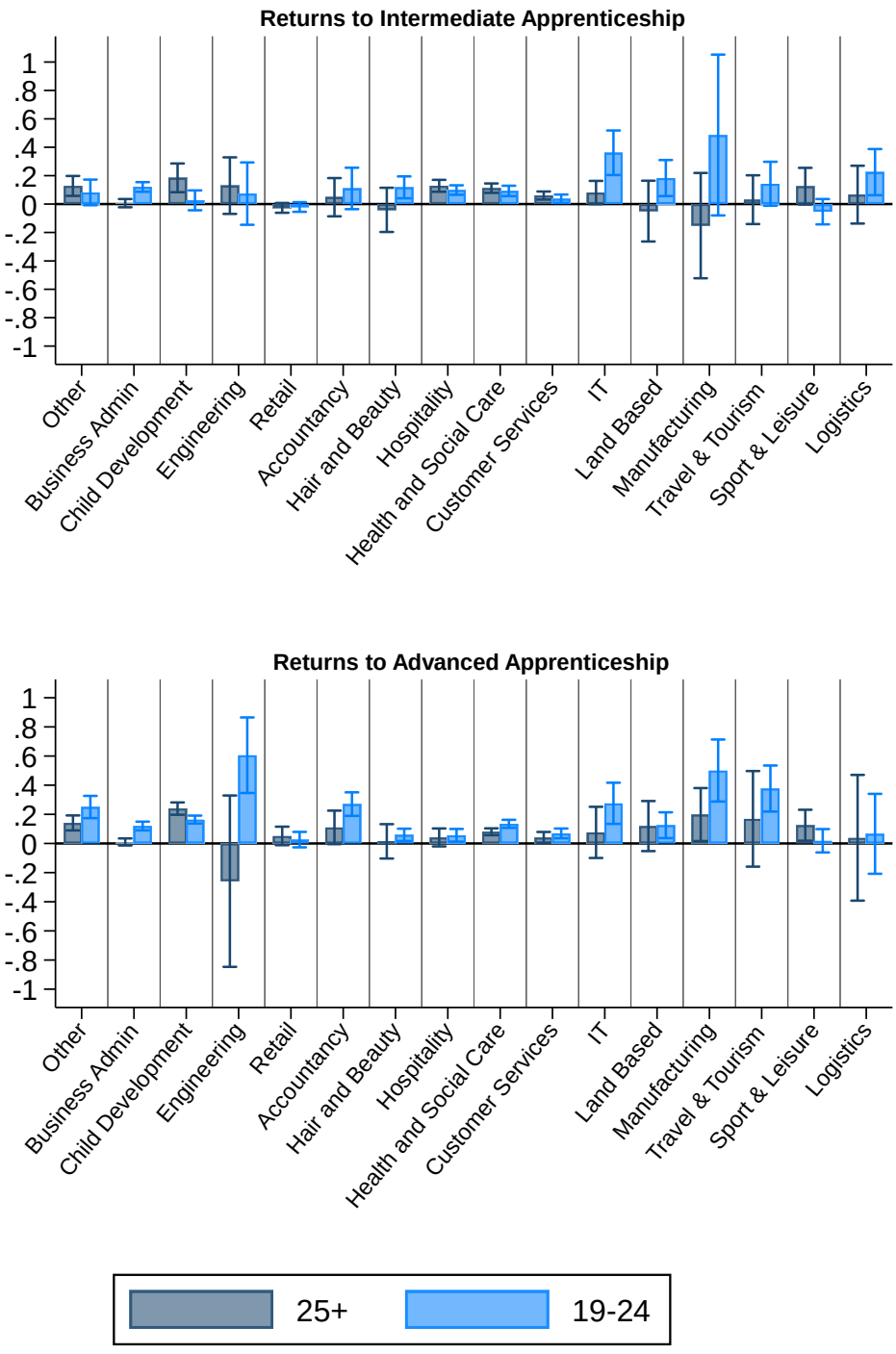


Figure 5: Apprenticeship Returns by Framework - Females



Note
Frameworks omitted due to large standard errors - Construction, Plumbing, Automotive, Electrotechnical

Table 1: Difference-in-Difference Estimates of Apprenticeship Earnings Differentials – Restricted Sample

	Intermediate Apprenticeship				Advanced Apprenticeship			
	Male		Female		Male		Female	
	25+	19-24	25+	19-24	25+	19-24	25+	19-24
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
3 Year windows								
Diff-in-Diff	0.066**	0.141**	0.057**	0.122**	0.075**	0.204**	0.052**	0.114**
	(0.021)	(0.016)	(0.017)	(0.017)	(0.020)	(0.015)	(0.015)	(0.014)
N	18570	43218	30450	37518	16362	51822	37968	53118
2 Year windows								
Diff-in-Diff	0.089**	0.115**	0.065**	0.085**	0.052**	0.198**	0.060**	0.126**
	(0.015)	(0.012)	(0.012)	(0.013)	(0.016)	(0.012)	(0.012)	(0.011)
N	39196	83772	58784	71028	30548	95432	64564	93936
1 Year window								
Diff-in-Diff	0.090**	0.135**	0.047**	0.096**	0.073**	0.201**	0.078**	0.144**
	(0.012)	(0.012)	(0.010)	(0.013)	(0.015)	(0.012)	(0.011)	(0.011)
N	66548	93764	97792	81522	38288	95706	83572	97710

Standard errors in parentheses, * p < 0.05, **p < 0.01. All regressions estimated using OLS. Reported coefficient is that on the interaction between apprenticeship completion (treatment) and a post-apprenticeship time period indicator. All regressions control for duration of apprenticeship, age completed, ethnicity, and current tax year dummies.

Table 2: Difference-in-Difference Estimates – Unrestricted Sample

	Diff-in-Diff - unrestricted sample							
	Intermediate Apprenticeship				Advanced Apprenticeship			
	Male		Female		Male		Female	
	25+	19-24	25+	19-24	25+	19-24	25+	19-24
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
3 Year windows								
Diff-in-Diff	0.052**	0.138**	0.025**	0.082**	0.081**	0.206**	0.093**	0.134**
	(0.008)	(0.007)	(0.007)	(0.007)	(0.009)	(0.007)	(0.007)	(0.006)
N	279987	384691	363028	336294	172545	390066	371711	403266
2 Year windows								
Diff-in-Diff	0.056**	0.145**	0.037**	0.092**	0.078**	0.214**	0.100**	0.142**
	(0.008)	(0.007)	(0.007)	(0.008)	(0.010)	(0.007)	(0.007)	(0.007)
N	207169	286041	272259	249898	126176	287976	272897	297068
1 Year window								
Diff-in-Diff	0.056**	0.147**	0.039**	0.098**	0.075**	0.213**	0.103**	0.147**
	(0.010)	(0.010)	(0.008)	(0.010)	(0.012)	(0.010)	(0.008)	(0.009)
N	119654	159243	159819	139858	70377	157335	153647	163383

Standard errors in parentheses, * p < 0.05, **p < 0.01. All regressions estimated using OLS. Reported coefficient is that on the interaction between apprenticeship completion (treatment) and a post-apprenticeship time period indicator. All regressions control for duration of apprenticeship, age completed, ethnicity, and current tax year dummies.

Table 3: Difference-in-Difference Estimates – Unrestricted Sample – Restricted Age Groups

	Diff-in-Diff - unrestricted sample							
	Intermediate Apprenticeship				Advanced Apprenticeship			
	Male		Female		Male		Female	
	25-29	20-24	25-29	20-24	25-29	20-24	25-29	20-24
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
3 Year windows								
Diff-in-Diff	0.087**	0.105**	0.053**	0.067**	0.096**	0.175**	0.077**	0.105**
	(0.017)	(0.007)	(0.018)	(0.008)	(0.017)	(0.007)	(0.015)	(0.007)
N	60318	292424	50921	254633	49980	277356	70245	317172
2 Year windows								
Diff-in-Diff	0.100**	0.111**	0.069**	0.076**	0.103**	0.185**	0.079**	0.112**
	(0.017)	(0.008)	(0.019)	(0.009)	(0.018)	(0.008)	(0.016)	(0.007)
N	44901	215234	38211	187692	36629	201998	51496	231367
1 Year window								
Diff-in-Diff	0.096**	0.116**	0.082**	0.083**	0.094**	0.186**	0.095**	0.120**
	(0.021)	(0.011)	(0.023)	(0.011)	(0.022)	(0.011)	(0.019)	(0.010)
N	25728	119010	22368	104525	20316	109471	28944	126293

Standard errors in parentheses, * p < 0.05, **p < 0.01. All regressions estimated using OLS. Reported coefficient is that on the interaction between apprenticeship completion (treatment) and a post-apprenticeship time period indicator. All regressions control for duration of apprenticeship, age completed, ethnicity, and current tax year dummies.

Table 4: Difference-in-Difference Estimates – Restricted Sample – Placebo Treatments

	Intermediate Apprenticeship				Advanced Apprenticeship			
	Male		Female		Male		Female	
	25+	19-24	25+	19-24	25+	19-24	25+	19-24
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
3 Years Windows (Placebo)								
Diff-in-Diff	-0.040	0.016	0.041	0.056*	0.003	0.022	-0.038	0.065**
	(0.036)	(0.027)	(0.028)	(0.029)	(0.033)	(0.025)	(0.024)	(0.023)
N	9285	21609	15225	18759	8181	25911	18984	26559
2 Years Windows (Placebo)								
Diff-in-Diff	0.004	0.006	-0.028	0.017	-0.009	0.020	-0.015	0.031
	(0.024)	(0.020)	(0.019)	(0.020)	(0.025)	(0.018)	(0.018)	(0.017)
N	19598	41886	29392	35514	15274	47716	32282	46968

Difference-in-differences effects estimated around a 'fake' (placebo) engagement with apprenticeships between 1 and 2 years before the individual actually engaged.

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$. All regressions estimated using OLS. Reported coefficient is that on the interaction between apprenticeship completion (treatment) and the 'fake' (placebo) post-apprenticeship time period indicator. All regressions control for duration of apprenticeship, age completed, ethnicity, and current tax year dummies.

Table 5: Propensity Score Matching Treatment Effect Estimates

	Intermediate Apprenticeship				Advanced Apprenticeship			
	Male		Female		Male		Female	
	25+	19-24	25+	19-24	25+	19-24	25+	19-24
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ATE	0.125**	0.213**	0.075**	0.144**	0.109**	0.210**	0.082**	0.180**
	(0.017)	(0.012)	(0.014)	(0.015)	(0.020)	(0.014)	(0.012)	(0.013)
ATT	0.146**	0.186**	0.073**	0.150**	0.119**	0.212**	0.102**	0.175**
	(0.020)	(0.015)	(0.017)	(0.018)	(0.025)	(0.016)	(0.015)	(0.016)
ATU	0.087**	0.255**	0.079**	0.136**	0.094**	0.206**	0.059**	0.187**
	(0.022)	(0.014)	(0.020)	(0.017)	(0.027)	(0.016)	(0.016)	(0.014)
N	21591	35636	28451	28699	10805	24456	22733	26700
Rubins B	9.038	5.889	8.116	10.131	10.384	5.274	8.034	12.838
Rubins R	1.057	1.026	1.061	0.962	0.901	1.074	0.954	0.891
Unmatched								
% Median Bias	5.954	4.968	14.526	7.701	19.132	6.707	11.212	4.076
% Mean Bias	22.871	16.664	29.791	19.792	31.741	15.415	30.725	15.956
Matched								
% Median Bias	1.340	2.367	1.769	2.367	2.119	1.145	0.981	0.926
% Mean Bias	2.833	2.350	2.594	2.715	2.823	1.526	2.067	2.552

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$.

ATE: average treatment effect

ATT: average treatment effect on the treated (apprenticeship completers)

ATU: average treatment effect on the untreated.

Rubin's B statistic is the absolute standardised difference of the means of the linear index of the propensity score in the matched treated and non-treated groups.

Rubin's R statistic is the ratio of the variances of the propensity score index in the matched treated and non-treated groups.

Mean/median % standardised bias is the average % standardised bias across all explanatory variables, where the % standardised bias is the percentage difference between the sample means of an explanatory variable in the treated and non-treated samples.

Matching variables: apprenticeship duration, the year the apprenticeship was completed, the age at which the apprenticeship was completed, ethnicity and labour market outcomes prior to the apprenticeship (total real annual pay, total days employed, and total days in receipt of benefits in the year prior to beginning the apprenticeship).

Table 6: Log Daily Earnings Differentials – Level Below vs Non-Achievers Counterfactual

	Male				Female			
	Level Below	Non-achievers	Level Below	Non-achievers	Level Below	Non-achievers	Level Below	Non-achievers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Advanced App'ship	0.429**	0.267**			0.304**	0.150**		
	(0.004)	(0.006)			(0.006)	(0.007)		
Intermed. App'ship			0.401**	0.192**			0.262**	0.150**
			(0.007)	(0.006)			(0.010)	(0.008)
N	307203	82306	120752	79344	205430	56051	59124	52116

Standard errors in parentheses, * p < 0.05, ** p < 0.01. All regressions estimated using OLS and include unreported ethnicity dummies, cohort fixed effects, and a variable controlling for the time elapsed since the highest qualification was completed.

Table 7: Log Daily Earnings Differentials – Level Below vs Non-Achievers Counterfactual, Including NPD Controls

	Male				Female			
	Level Below	Non-achievers	Level Below	Non-achievers	Level Below	Non-achievers	Level Below	Non-achievers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Advanced App'ship	0.296**	0.223**			0.110**	0.107**		
	(0.004)	(0.007)			(0.006)	(0.008)		
Intermed. App'ship			0.216**	0.150**			0.082**	0.118**
			(0.009)	(0.007)			(0.013)	(0.009)
N	307203	82306	120752	79344	205430	56051	59124	52116

Standard errors in parentheses, * p < 0.05, ** p < 0.01. All regressions estimated using OLS and include unreported ethnicity dummies, local education authority fixed effects, cohort fixed effects, Key Stage 4 school fixed effects, GCSE attainment (A*-C in Maths. A*-C in English, number of GCSEs, number of A*-C GCSEs, KS3 Maths score, KS3 English score, KS2 Maths score, KS2 English score, free school meals entitlement indicator, and a variable controlling for the time elapsed since the highest qualification was completed.

Table 8: Decomposition of Age Group Difference in Earnings Differentials Within and Between Frameworks

	Percentage within frameworks	Percentage between frameworks
Male Intermediate	79.6%	20.4%
Female Intermediate	77.5%	22.5%
Male Advanced	35.7%	64.3%
Female Advanced	88.9%	11.1%

Table reports the proportion of the difference in earnings differentials between 19-24 year old and age 25+ apprentices that occurs within and between apprenticeship frameworks according to the formula $\Delta D = \sum_k \Delta D_k \overline{S_k} + \sum_k \Delta S_k \overline{D_k}$ where the first term is the within framework differences and the second term is the between framework differences.