UNIVERSITY OF Southampton
Southampton Statistical Sciences Research Institute

Working Paper A10/05
Applications and Policy

Who Has Realistic Income Expectations: Students Or Workers?

John Jerrim

Abstract

I analyse a nationally representative, longitudinal dataset to investigate the income expectations of 20 American year old men over a 10 year time horizon. Specifically, I compare the accuracy of students' expectations to young adults from the same cohort who have chosen to enter the labour force. The results indicate that 20 year old American men overestimate their income at age 30 by an average of 50%. Yet, there is little evidence that those in the labour force are more realistic than their peers in higher education. In fact certain groups of students appear to be more realistic than those who are working, overestimating their future income by less than 10%.

Who has realistic income expectations: Students or workers?

John Jerrim

Institute of Education, University of London

Department of Social Statistics, University of Southampton

Abstract

I analyse a nationally representative, longitudinal dataset to investigate the income expectations of 20 American year old men over a 10 year time horizon. Specifically, I compare the accuracy of students' expectations to young adults from the same cohort who have chosen to enter the labour force. The results indicate that 20 year old American men overestimate their income at age 30 by an average of 50%. Yet, there is little evidence that those in the labour force are more realistic than their peers in higher education. In fact certain groups of students appear to be more realistic than those who are working, overestimating their future income by less than 10%.

Key Words: Expectations, Students, Higher Education, National Education Longitudinal Survey

JEL Codes: I2 Education, D84 Expectations

Acknowledgements:
I would like to thank the feedback I received on this piece of work from members of staff at the University of Southampton (Dave Holmes, Sylke Schnepf), Institute of Education (John Micklewright) and University of Madison-Wisconsin (Karl Scholz). I would also like to thank those who commented on this work at the 2010 ESPE conference and during seminar presentations in Southampton and Madison. The usual disclaimer applies.

Contact Details: John Jerrim (J.Jerrim@ioe.ac.uk) Department of Quantitative Social Science, Institute of Education, University of London, 20 Bedford Way London, WC1H 0AL

"Many have argued that attitudes of investors in human capital are very different from those of investors of physical capital because the former tend to be younger, and young persons are especially prone to overestimate their ability and chance of good fortune"

Becker (1993) Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (Third Addition), page 93

1. Introduction

Economic models of schooling choice are often based on the assumption that young adults have a realistic idea of what their future income will be. This assumption can be decomposed into two parts. The first is that young adults hold "full information" – that they understand the economic benefits of the different educational options available. The second is that they hold "rational expectations"; that they are able to use this information efficiently to produce realistic assessments of the future. If this is true for all members of the population, then there should be no systematic differences (on average) between individuals' apriori expectations and later realisations.

It is under these assumptions that economists often use realised ("ex-post") income data to estimate how young adults' expected financial returns to education influence their schooling choices. Notable examples include Berger (1988), Willis and Rosen (1979) and Boudarbat (2004). They all find that expected returns have a large and statistically significant effect on young adults' decisions, whether this is to enter higher education at all or the specific subject they take. However, these results are heavily reliant on the strong assumptions described above. If these are violated, then one may question the robustness of such results.

In Jerrim (2008) I illustrated that UK students, on average, overestimate their starting salary. Yet I also found that the accuracy of labour market predictions varies substantially between different groups. In particular, those who probably held the least labour market information were, on average, the least realistic. This has raised some question marks over whether the assumptions described above hold. However, this analysis was limited by the fact I only considered the accuracy of students' expectations in a single setting over a short time horizon. From an economic perspective, it is lifetime income, rather than starting salaries, that influences people's decisions. An interesting extension of the work completed in Jerrim (2008) is therefore whether students in other settings, and over longer time horizons, are just as unrealistic about their future income.

Moreover in Jerrim (2008) I only considered the expectations of young adults in higher education. Indeed, all other studies, that I am aware of, have done the same; those who have chosen to enter the labour market straight from school have typically been ignored. Yet this group is both relatively large in size and of substantial interest. In particular, workers may have the opportunity to collect valuable information about the labour market from their employers, colleagues and the job search process itself. A student trying to access the same information may face much higher costs. If both groups are “rational”, processing all the information that they hold efficiently, one would expect workers’ additional information to translate into more accurate expectations. On the contrary, one may argue that students are less myopic than their peers who enter the labour force, and thus less readily discount their future income. Consequently, they may have more incentive to collect information about long-term labour market outcomes and hence hold more realistic expectations. Likewise, it maybe that children who invest in higher education are focused on one particular career, while 20 year olds in the labour force perhaps move somewhat haphazardly between different types of job. Thus it may be that students seek out more specific and relevant information than their working peers, leading to a more realistic assessment of future labour market opportunities.

Indeed this argument may hold true for some groups of students, but not for others. Those studying Education, Nursing or Engineering are being trained for specific jobs. With career councillors widely available in almost all Higher Education institutions, these students are probably well informed about the graduate labour market. On the other hand, someone in a course not leading to one particular career, such as Arts, Humanities, Languages or Social Sciences, may only have a vague idea about the type of job they will pursue. Thus these students receive only quite broad, low quality labour market information and will thus be prone to either under or overestimation of their future income. The discussion above illustrates the interesting insights that a comparison between different groups of students and workers might bring.

At present, such topics have received very little attention in either America or Europe. There are some small scale US studies that compare *students*' expectations to actual labour market outcomes. However, these suffer methodological difficulties, and results can not be generalised to the wider student population. Furthermore students are generally treated as a homogeneous group. There is little discussion of the association between dropping out of college, idiosyncratic ability, subject being studied and the accuracy of future expectations. Of particular note, no comparison is made to the expectations of their peers in the labour force. I make a significant contribution to the small US literature by using a detailed, nationally representative sample of both students and young workers to consider how the factors listed above influence the accuracy of 20 year old men's income expectations ten years into the future.

My results suggest that the US student population over-estimate their future income. However, unlike existing studies, I show that this result only holds for certain groups; I find that some students actually make quite good long range predictions, overestimating their income ten years into the future by (on average) less than 10%. Moreover, I find that differences between students and workers are not as pronounced as one may expect; under certain conditions, students actually hold more realistic expectations than their peers in the labour force.

I begin in section 2 by reviewing the current literature on income expectations and motivating the need for this research. In sections 3 and 4 I describe the National Education Longitudinal Survey (NELS) data. This is followed in section 5 by my analysis of young adults' labour market expectations. I conclude with a discussion of my key findings, and argue that either young adults do not hold enough labour market information to predict their income at age 30, or simply choose not to incorporate it into their expectations.

2. Literature and research questions

To my knowledge, there are six published US studies that investigate *students'* income expectations. The first to consider this topic was Smith and Powell (1990). They asked 400 students at two mid-western universities how much they expect to earn when they graduate and after 10 years in the labour market. Respondents were quite realistic about pay in their first job, but overestimate wages in 10 years time. Betts (1996) asked 1,000 students at the University of California to predict wages for a hypothetical individual under several different scenarios. He finds that students quite accurately predict the wages of young workers, but overestimate the pay of those with ten or more year's labour market experience. Blau and Ferber (1991) collected data from 351 students studying in the Business faculty at the University of Illinois. Again, students seem quite realistic about starting wages, but become progressively unrealistic over long time horizons. Carvajal et al (2000) analyse the expected starting salary of 219 Business students at Florida International University. They find over-estimation of around 10%. Rouse (2004) investigates the wage expectations of 69 high school seniors from the Baltimore City Public School District[1]. She finds these high school seniors to be quite unrealistic about their future income at age 30. Dominitz and Manski (1996) take a different approach. They asked 110 Madison students several questions to try and not only capture individuals' expectations, but also their uncertainty about future outcomes. They find that male students are reasonably realistic, but girls less so.

These studies suggest that students have a reasonably good understanding of starting salaries, but are less realistic about their future income over longer time horizons. However, this relatively small literature is somewhat limited by the scope and design of the aforementioned studies. Data are typically:

(a) Collected from students at just one (or at most two) universities
(b) Drawn via convenience sampling, rather than a probabilistic method
(c) Over-represent students from mathematical subjects (Economics, Finance, Engineering) and under represent those studying Art
(d) Of limited sample size

[1] Rouse (2004) also uses the NELS data analysed in this paper. In particular, she compares the NELS expected income data to external estimates on actual wages drawn from the 1990 Census. I analyse the NELS data in greater depth than Rouse, and focus on a set of quite different hypotheses. In particular, she is concerned with differences between ethnic groups, where my concern is the accuracy of students expectations compared to workers.

This causes several methodological problems. Firstly, as samples are often drawn from one university and a handful of subjects, it is difficult to generalise results to the wider US student population. This leads to problems when the authors try to assess whether students hold "realistic" expectations. Wage expectations, drawn from a highly selective survey, are compared to data on national graduate wages from an external data source, such as the Current Population Survey (CPS) or Census. The wage expectation and realisation data often represent two populations that could differ in all manner of characteristics. This may clearly bias any assessment of whether expectations are realistic. Secondly, small sample sizes mean that the wage expectation data suffers from quite large sampling error. But, as data are usually drawn via convenience sampling, reliable standard errors can not be calculated. Hence the true extent of sampling error on results actually often goes unknown.

In addressing my first research question, I attempt to overcome these limitations by analysing the income expectations of a large, nationally representative sample of American adults using NELS 1988 data. As this survey was designed to be nationally representative, my results should generalise to the wider US student population. Furthermore, as expectations and realisations are collected from the same individual over time (i.e. this is a panel dataset), my results should also be driven less by the composition and selectivity of the sample than the small scale studies cited above. Moreover, as the NELS data was collected using a large, probability based sample design I can adequately demonstrate the influence that sampling variation has on my estimates. In summary, my first research question is:

Q1. Do 20 year old male students in the US, on average, have realistic expectations of their income at age 30?

However students in countries like the USA, with its large and diversified higher education system, are not a homogeneous group. The accuracy of their labour market expectations is probably related to the subject they study, whether they actually graduate from university and their underlying cognitive ability. For instance, young adults who begin, but do not complete, university probably form their expectations based on the belief that they will obtain a degree. They may not adequately account for the possibility of dropping out, and hence (ex-post) their expectations will appear overly ambitious[2]. Similarly, given the results I found for the UK in Jerrim (2008), one may suspect the accuracy of students' expectations to vary

[2] This is something that I shall go on to explore in section 5.

substantially with the subject they study. Indeed, as noted in the introduction, students studying certain subjects may hold more information about the labour market than others. This may be because they are already being trained to enter a specific job (teacher, nurse, engineer) and hence collect specific and detailed information compared to their peers entering the more general graduate labour market. Alternatively, it may be that wages within these jobs, or the wages of previous graduates from similar disciplines, have quite low variability. Hence students within these subjects face less uncertainty than some of their peers. In either case, one might expect to find similar patterns to those observed in the UK, where students studying mathematical and vocational subjects make better predictions than their peers in more creative disciplines. Such details have rarely been discussed in the US literature, hence my second research question:

Q2. Do students in the US who drop out of university hold particularly unrealistic labour market expectations? Are maths and vocational students more realistic about their future income than those studying more creative subjects?

As the reader may have noted, all the studies reviewed focus on students at university. I do not know of any work that investigates the expectations of young adults who have chosen to enter the labour market rather than continue their education[3]. Apriori, one may expect young workers to hold more realistic expectations than students, as they probably have greater access to relevant labour market information. For instance, they will have contact with older workers who, either formally or informally, pass on sector-specific details of future pay and progression. Alternatively, organisations themselves could make information on career progression and pay freely available to their staff. Another factor is that workers have been through the job search process at least once. They should have found out about wages and career opportunities during this time. Indeed, these individuals may have held unrealistic expectations before this experience, but actually going out and trying to find a job may have taught them the realities of their employment opportunities[4]. Many university students would not have had a similar experience of searching for a full-time job, and may still be holding

[3] Dominitz (1998) assesses the accuracy of American workers' wage expectations. However he does not specifically investigate the expectations of young workers, or how realistic they are compared to students. He also focuses on wage expectations for the year ahead, whereas this paper looks over a longer time horizon.

[4] Recent work in the sociological literature by Morgan (2005) depicts young adults as "Bayesian learners". In particular, he illustrates the accuracy of a "fast" and "slow" learner's expectations over time. Morgan suggests the difference between fast and slow learners could be to do with the different timing of key life events. This could include entry into the labour market, a period when young adults should receive a lot of information that will lead them to quickly (and more accurately) updating their expectations.

onto their unrealistic expectations. Self-selection into the labour market or higher education may also play a role. Educational attainment is linked to migration (see Borjas 1999). The less educated (who have self-selected into work) are more likely than students to stay in their age 20 location. Workers therefore gather information about wages in the local labour market that they incorporate into their expectations. Students, on the other hand, may well expect to work in other areas of the country. The local labour market will be less informative for many of them. Hence one would expect the labour market to have more salience to those who are already actively employed[5].

Based on these discussions, my third and fourth (inter-related) research questions are:

Q3. Can 20 year old US workers, on average, make realistic predictions of their income at age 30?

Q4. Do 20 year old US workers, on average, make better predictions of their income at age 30 than their peers in higher education?

Although these hypotheses are similar, they do pertain to slightly different things. In particular, it is possible that both workers and students overestimate their future income (reject the null hypothesis of no difference between expectations and realisations in Q1 and Q3) but for workers to still make better predictions than students (reject the null hypothesis that students and workers are equally unrealistic in their labour market expectations in Q4).

This work adds significantly to the literature reviewed at the start of this paper. I know of no other study that analyses nationally representative data on young adults' income expectations. Moreover, to my knowledge, I am also the first author to use panel data to compare income expectations and realisations over a 10 year time horizon. Thus I can more accurately compare students' expectations to later realisations, with my results being more likely to generalise to the wider US population. I am also able to tackle several new and interesting hypotheses that put the existing work on students' expectations into a wider context.

[5] Counter-arguments to this hypothesis have been presented in the introduction to this paper.

3. Data

To address these research questions, a nationally representative data source is required that follows young adults from their initial predictions of future income to their later success in the labour market. This needs to follow an entire cohort of young adults and not just those who continue on to university. One source is the National Educational Longitudinal Survey (NELS) from the US. This study's aim was to provide data about adolescents at critical points in their development and later into their careers using a nationally representative sample of adolescents. Children were initially interviewed in 1988 when the majority were 14 years of age. They were then followed up four further times, at ages 16, 18, 20 and 26. Parents and teachers of the pupils also completed the first three rounds of the survey.

In the first wave (age 14), a two-stage stratified sampling design was employed, with schools as the primary sampling unit, and probability of selection (of schools) proportional to size. 1,052 schools participated in the survey, including some oversampling of private institutions. A random sample of 26 students was then selected from each school. 26,432 students were eventually selected with 24,599 taking part (93%). In each of the next two waves (age 16 and 18) students who participated in the initial survey were followed up. The sampling process added some newly selected students (1,043 at age 16 and 244 at age 18) [6]. This was done to create a valid probability sample (a nationally representative cross-section) of students in each of the respective years. In total, 20,923 18 years olds took part in the third wave. The fourth wave took place when students were 20 years old. To reduce costs, a sub-sample was selected based on demographic characteristics and response history. It is important to note that this reduction is not the result of sample attrition, but from a conscious effort of the survey design to limit burden and cost[7]. This led to the age 20 sampling frame being reduced to 15,964 individuals. In total, completed responses were available from 14,915 (93%) 20 year olds. Further details are given in Appendix 1. The final survey took place when most sample members were 26. Data are available for 12,144 individuals (76% of the age 20 sub-sample).

[6] These students were not randomly selected, but drawn from schools where there were other second and third wave respondents. More details can be found in Appendix 1 and page 56 of Curtin et al (2002).

[7] Around 5,000 individuals were dropped from the study. 2,000 of these were classed as "poor responders", who were basically excluded because of the low chance of future contact. Hence it may be more appropriate to consider these 2,000 observations as non-respondents. The other 3,000 individuals dropped were not classified as poor responders, but excluded purely to lower costs. Further details are given in Appendix 1.

There are obviously some issues of non-response due to sample attrition. One way to help correct for differential response rates in terms of observable characteristics is the use of survey weights. The NELS dataset contains a cross-sectional weight for those who took part in the final survey, and various panel weights. Unfortunately a panel weight is not provided for those who completed the final *two* surveys (ages 20 and 26). Instead a panel weight is available for those who had completed the final *three* surveys (ages 18, 20 and 26)[8]. This refers to the population of high school seniors in 1992. The National Centre for Education Statistics describes this panel weight:

This is the second, third, and fourth follow-up panel weight, which applies to the 12th grade cohort. It applies to fourth follow-up respondents (i.e. 2000) who were also respondents in the second and third follow-up rounds (i.e., 1992, 1994). It estimates longitudinal parameters that describe the population of spring 1992 12th graders.

This weight shall be used in all subsequent analyses to help adjust for unit non response and over sampling of certain minority groups. Therefore, the population I am describing in this analysis is those who were high school seniors in 1992[9].

A vital question is how respondents were asked to report their future income expectations. When respondents were 20 years old, they were asked:

"What do you expect your total annual income to be when you are 30 years old?"

This question is comparable to those asked in the other major studies of students' labour market expectations[10]. Respondents are clearly asked to predict *their* future income. However this raises several issues about how people actually respond to this type of question. Do they take into account inflation? Do they report gross or net earnings? Is this conditional on having a full-time job? To shed some light on these issues, it is important to consider the

---

[8] Around 98% of those who responded at 26 also responded at ages 18 and 20.

[9] Some young adults drop out of school before this point (senior high school year). Consequently, I may not be representing this quite small sector of the US population. This is further highlighted where I compare the NELS sample to CPS data in Appendix 2.

[10] For instance, Webbink and Hartog (2004) phrase their question "How much will your net starting salary be after graduation?" Betts (1996) asks students the question "Below, please circle your estimate of the national average for annual starting salaries". Bruenllo et al (2001) ask "What do you expect to earn right after finishing your degree (first degree possible at your university). State an approximate amount per month (net, i.e. after paying taxes)?"

ordering of survey questions. Indeed *directly before* they were asked for their income expectations, respondents were given the following question:

*"What was **YOUR** total income from all sources, before taxes, in 1993?* [i.e. the year prior to interview] *This figure should include salaries, wages, pensions, dividends, interest, unemployment compensation, grants, financial aid, scholarships, government assistance (AFDC) and all other income"* [Capitalisation in original question]
Write in dollar amount, write in 0 if no income $..

I assume that respondents follow the same criteria to the above when reporting their income expectations. For example, I assume respondents use their answer to the question above as a reference point for their income expectation and answer with current prices in mind. Indeed, this assumption is consistent with the existing literature. Therefore all reported expectations are assumed to be in 1994 prices. It is also quite clear that respondents should be reporting gross (pre-tax) figures. The question also asks for total annual *income*. This would suggest respondents should not only take into account future wages, but other sources of income such as receipts from benefits or interest payments when reporting their expectations. Another point to note is that the question on income expectation asks for the respondent's ("your") expected income. This is made even clearer in the preceding question, with **"your"** in bold, capital letters. It seems that the respondent should only be considering their own, personal income, and not their partner's or other family members'.

Dominitz and Manski (1994) state that both men and women respond to questions on income expectations conditional on holding a full-time job. This seems a reasonable assumption for male respondents. However women may expect to have children and be out of the labour force, or working part-time, by the time they are 30. Alternatively women may report their expectation based on working full-time, but by the time they are 30 and have a child, self-select out of the labour market or into part-time jobs. Hence any comparison of women's expectations to later realisations has additional complications. In particular, I do not know how or whether women incorporate selectivity out of the labour force (due to childbearing) into their expectations. I could proceed by assuming that their reported figures are conditional on working full-time (as suggested by Dominitz and Manski). Yet even then I would face the non-trivial task of trying to control for this self-selection in the observed income data. Thus, although I note the potential interest in this issue, and that the problems discussed above are perhaps solvable, I focus on only the 5,782 male responses in this analysis for brevity.

From the initial sample of 5,782 male observations I exclude 633 respondents with missing expectations data from the analysis. A further 39 observations are excluded when an individual reported expected income below $6,000. On the assumption of full-time working, stated above, figures below this level would violate minimum wage laws in the USA. Another 71 observations were dropped where expectations were over $250,000. In total, 743 (13%) male observations have been excluded due to difficulties with the income expectations data.

I investigate the characteristics of those excluded with a logistic regression of item non-response. The results, presented as odds ratios, can be found in Table 1.

TABLE 1 about here

It seems that respondents from the wealthiest backgrounds are the most likely to report their expected income. Similarly, young adults of American Indian descent are less likely to report their expected income than Whites. Of importance for my substantive research questions, it appears that 20 year old students are more likely to respond than their peers in the labour force. More reassuringly, there is no association between wages recorded at age 26 and missing expectations data. In other words, it is *not* the case that those who reported particularly high or low wages at age 26 are the individuals who did not report their salary expectations[11]. Likewise, it does not seem to be the case that missing expectations data is related to the respondents' maths ability or whether they are a student who drops out of university before they complete their degree[12].

Nevertheless Table 1 does indicate some self selection into the study. If those who choose to take part are more (or less) realistic than those who do not, I would underestimate (or overestimate) the difference between average expectations and later realisations. Likewise, the fact that workers are more selective about taking part than students could introduce bias

[11] I also tested for an interaction between wage at 26, and whether the respondent was a student or worker at age 20. The coefficient was neither big nor statistically significant.
[12] In a specification not presented, I entered students in groups depending on their subject area. I found no statistically significant differences

when I consider which of these groups are more realistic (research question 4). For instance, only the most optimistic workers may report their expected income. On the other hand both optimistic *and* cautious students may offer a response. In this scenario, my selection of workers would appear to be less realistic than a true random sample from the population. I have checked the robustness of the results I present in section 5 to this non-response by creating and applying a set of response weights, with estimates presented later in the paper[13].

For those individuals with complete expectations data, Figure1 presents the distribution of expected age 30 income. Notice firstly the large, positive skew of the data, with a mean (median) of $50,312 ($40,000) and standard deviation of $30,051. This result is driven by a number of large observations; the top percentile expect to earn over $200,000 per year at age 30 (in 1994 prices). One may ask whether this variable is truly reflecting individuals expectations (what they *think* will happen) rather than aspirations (what they *hope* will happen). In this paper, I go on to assume the former, but one can not rule out some individuals adhering to the latter. I have excluded some very large observations, where the figures maybe reflecting children's "aspirations" rather than their "expectations". In section 5, I present a set of robustness checks using quantile (median) regression to see how results differ when I do not exclude these data.

FIGURE 1 about here

A second feature of the distribution is the bunching of observations. Over half of all observations lie at five points: $30,000 (13%), $40,000 (14%), $50,000 (14%), $60,000 (7%) and $100,000 (3%). I describe a similar phenomenon for UK students in Jerrim (2008). The general explanation is that respondents are rounding their responses to the nearest $5,000 or $10,000. This may reflect uncertainty about future income or, on the other hand, that individuals simply think in terms of round numbers. The implication is that there is some rounding error in individuals reported expectations. When considering expectations at the group or population level, it seems reasonable to assume that this rounding error will be on

[13] Note that the effectiveness of such weights in correcting bias is dependent upon the explanatory power of the underlying non-response model. Table 1 indicates that few of the covariates included in the logistic regression are statistically significant. Consequently, one may expect the results to appear no different as model used to create the response weights is relatively weak.

average zero (some individuals round up, others round down) [14]. On the other hand, considering expectations at the individual level, this is a potentially important source of measurement error.

The NELS also contains data on each respondents' wage history. As part of the survey at age 26, they were asked:

For your (current/most recent) job, about how much (do/did) you earn before taxes and other deductions?

Enter Amount $..

Interviewer instruction: Record the time scale of the amount (e.g. $30,000 per year)

1 = hourly, 2 = weekly, 3 = every two weeks, 4 = monthly, 5 = annually

How many hours per week, in a typical week (do/did) you (currently) work for pay in your job as a/an [F4BOCTV (verbatim job-title)]?

Enter hours per week:………………………Hours

This clearly asks for gross earnings in their current or most recent job. However the respondent could choose the timescale to report this figure, with a breakdown provided in Table 2[15].

TABLE 2 about here

For those who provided a weekly, fortnightly or monthly figure, I have scaled their pay up to the annual equivalent. All respondents were also asked how many hours they work in an average week. For those reporting an hourly wage, this was used to calculate their annual equivalent.

Wages from previous years were also collected retrospectively at age 26 (the final survey

[14] Obviously this assumes respondents round their expectations as would a mathematician, rather than another rule (for example, always rounding up to the next highest multiple of $5,000).

[15] Note the difficulties when recording salary details because of measurement error, with results based on surveys often different to that held in administrative records. Indeed this measurement error could vary by the unit of time respondents' answer in (see Cartenseen and Woltman 1979).

wave)[16]. Respondents were asked:

First, including all of the wages, salaries, and commissions you earned in (1997/1998/1999), about how much did you earn from employment before taxes and all other deductions?

Again gross wages are recorded, containing details on all forms of employment related income, including all commissions, tips and bonuses. Therefore, the NELS data has reported wages for respondents between the ages of 23 and 26[17,18].

Previously, I stated my assumption that respondents are providing their income expectation conditional on holding a full-time job. Thus I only consider realised wages when the respondent was working full-time[19]. Those with no history of full-time work have been excluded from the analysis. Table 3 shows that, by doing this, I exclude a further 605 (12%) observations.

Of course, like the missing expectations data, this may introduce some selectivity into the sample. Individuals may choose not to work full-time between the ages 23 and 26. An obvious example is graduate students, many of whom remain in education throughout their early twenties. If these individuals are substantially more or less realistic than other groups, then some selection bias may be introduced into my results. Alternatively, there could be a direct relationship between income expectations and selection out of work. Those with unrealistically high income expectations may also have unrealistically high reservation wages. These individuals are less likely to receive a suitable wage offer, and therefore choose not to work. In this situation, I would be excluding the most unrealistic individuals from the analysis. In Table 4 I present a logistic regression that investigates this possible selectivity.

TABLES 3 & 4 about here

---

[16] Measurement error due to recall bias poses a possible difficulty in using this data. See Bound et al (2001) for further details on the difficulty of recording historical wages with retrospective questions.
[17] I deflate all information on actual wages and unearned income into 1994 prices using data from the Annual Wage Index, available at http://www.ssa.gov/OACT/COLA/awidevelop.html
[18] For 80% of respondents, data on full-time wages is available for at least 3 of these 4 years.
[19] Where gaps appear in individuals wage profiles (between 23 and 26), information from previous years (when they were working full-time) shall be extrapolated forward to estimate age 30 wages. Further details follow in section 4 and Appendix 2.

Respondents who have parents in the top quartile of the income distribution are three times more likely to respond than their peers whose family income is in the bottom quartile. Similarly, those with reported health problems are more likely to be excluded than those without. On the other hand, it seems that respondents who were students at age 20 are just as likely to be excluded as those who were working. However, those who were unemployed at age 20 are relatively unlikely to have a full-time wage recorded between the ages of 23 and 26. Reassuringly, there is little evidence that those with the highest wage expectations were the least likely to be working full-time between the ages 23 and 26. Interestingly, specification 3 and 4 show that low ability respondents were *less* likely to be excluded from the analysis because of missing income data. It seems the brightest sample members tend to either not report their salary or have selected out of full-time work up to age 26 (e.g. to continue their education). If individuals of high ability are more efficient at processing labour market information and thus hold more realistic expectations, then their exclusion may have an influence on my results. The final specification shows that those who are still studying at age 26 are the most likely to be excluded. Further analysis not presented indicated that around half the excluded observations came from individuals who were studying full-time at age 26. It is likely that these individuals have never left higher education, and hence have no full-time wage history. The main implication seems to be that certain groups of students are likely to be excluded from the analysis; particularly those who continue onto graduate school and remain in education through their early 20's. If these individuals hold significantly more (or less) realistic expectations than other groups of students, then my results could again be influenced by their exclusion.

I further investigate for selection from missing data in Table 5 by presenting a set of summary statistics. The left hand column illustrates the characteristics of the initial 5,782 male observations in the complete NELS sample, while the column on the right shows the characteristics of the 4,434 individuals who are not missing any key information. Reassuringly, the distribution of observable characteristics remains reasonably similar.

TABLE 5 about here

Though the primary focus of this paper is young men's expected income, I present some additional results referring to other aspects of their anticipated labour market success as a robustness check. In particular, I put forward the argument that if students are unrealistic about their future income, they are also likely to be unrealistic about other aspects of the labour market, like their future occupation. Analogous to finding excessive income expectations, individuals may expect to be in a professional occupation when they turn 30, but actually end up working in a relatively low paying job. As part of the NELS survey at age 20, individuals were asked what occupation they thought they would be working in at age 30[20]. In the final survey wave (age 26) individuals were asked what occupation they currently hold. Therefore I also compare expected and realised occupation to support my main analysis surrounding young adults' income expectations.

At this point, however, one should note that there are two significant problems with comparing expectations and realisations using the NELS data:

(a) At age 20, respondents were asked what they expect their income (and occupation) to be at age 30. However, data on labour market realisations is only collected between the ages 23 and 26.

(b) Respondents are asked about their expected *income*. Data on realisations focuses on *wages*.

I go on to discuss these points in section 4 and Appendix 2. Specifically, these sections cover how I use the information available to predict individuals age 30 income. To conclude this section, I simply ignore such issues and compare expectations of *income* at age 30 to realised *wages* at age 26.

Figure 2 presents the distribution for age 26 (actual) wages. Comparing this to the expected income distribution at age 30 in Figure 1, there is significantly less bunching and positive skew. The standard deviation is much smaller ($16,479 for the actual age 26 wage distribution compared to $30,051 for the age 30 expected income distribution), though there is little difference in the decile ratio (3.4 compared to 3.2).

FIGURE 2 about here

[20] The exact wording of the question was: "What job do you expect or plan to have when you are 30 years old?" Respondents were asked to write in an occupational description into an open text field.

Further insights come from investigating the ratio of the 10th to 50th percentile (p10/p50) and the 90th to 50th percentile (p90/p50) . The bottom halves of the distributions (p10/p50) are very similar (0.6 in the actual distribution compared to 0.625 in the expected). The difference is slightly bigger in the top half of the distribution, with the p90/50 for actual wage (1.8) below that in the expectations (2.0). There is also some initial evidence that young adults' expectations may be somewhat optimistic. The median (mean) expected income at age 30, in 1994 prices, is $40,000 ($50,312). Comparatively, the median (mean) actual wage of 26 year olds stands at $23,079 ($26,210). For expectations to be realistic on average by age 30, I would need to find that either:

(a) average annual real wage growth is around 15% between the ages 26 and 30.
or
(b) 30 year old men, on average, have large quantities of unearned income.

I turn to these two topics in the following section.

4. Prediction of age 30 income

Respondents are asked what they expect their annual income to be when they turn 30, but realised wage data is only available between ages 23 and 26. In Appendix 2 I fully set out two methods of predicting age 30 income. In this section, I will provide the intuition behind these methods. Specifically, I separate this into two parts: (a) the estimation of wages, and (b) the estimation of unearned income.

Wages

Figure 3 illustrates the data observed for one particular individual in the NELS[21]. At age 26, this individual has a particularly large wage by his "historical" (age 23 - 25) standards. This may be a *permanent* shift in his wage profile, for instance a change in career. In this case, previous earnings have little relevance for predicting future wages. In contrast this could be a *temporary* shock to his wage, for instance a salesman who has had a particularly good year.

[21] For this particular individual, the income they expect is significantly higher than their wages recorded at age 26. This is not typical of all other respondents in the dataset. Rather I have chosen this individual as he is a good example of the substantial points I make throughout this section.

In future periods, his wage will revert to its historical average (i.e. the average of the previous 3 years). On the other hand, reality may lie somewhere in-between these two extremes, perhaps reflecting the fact that he happened to receive a large pay rise that year.

Given these possibilities, I use two methods to predict age 30 wages. The first method views large wage changes as a *permanent* shift in an individual's earnings profile. Under this method, I simply take the most recently observed wage (age 26) for each individual and extrapolate it forward (to age 30). Figure 4 presents a hypothetical example.

FIGURES 3 & 4 about here

To implement this method, I use estimates of the annual real wage growth for young workers provided by Rubinstein and Weiss (2007). Specifically they provide a table of average annual real wage growth rates broken down by labour market experience and educational attainment for three surveys; the Current Population Survey (CPS), Panel Survey of Income Dynamics (PSID) and National Longitudinal Survey of Youth 1979 (NLSY 79)[22]. The growth rates they calculate from the CPS, PSID and NLSY are provided in Table 6, with further details available on page 14 and Appendix 5 of Rubinstein and Weiss (2007).

TABLE 6 about here

These growth rates are applied to each individual in the NELS, depending on their highest qualification achieved by age 26[23]. From this point on, I shall call this prediction "Method 1". Note that, for all groups, average real wage growth rates are under 6% per annum. This is well below the 15% per annum that I suggested NELS sample members needed in the unobserved period (i.e. between ages 26 and 30) for their expectations to be (on average) "realistic" (recall my brief discussion at the end of section 3).

[22] Rubinstein and Weiss restrict each of the above datasets to full-time, male, American workers, as I have done with the NELS. One should note, however, that these surveys all relate to different years. The CPS data relates to wages between 1998 and 2002, the PSID is for all years after 1968, while the NLSY draws its information between 1979 and 2000.

[23] For example, an individual with college education, and who was earning $50,000 dollars at age 26, would be estimated to be earning $61,240 at age 30 (all in 1994 prices). This is calculated by $50,000 * ($1.052^4$), using the NLSY data and "College graduates" column in Table 6. In the event that wages go unobserved at age 26 (e.g. the individual was unemployed) I extrapolate from their last observed full-time wage. For example, if someone was earning $50,000 at age 25, and their wage was not recorded at age 26, I would predict their age 30 income to be $50,000 * ($1.052^5$) = $64,577

On the other hand, prediction "Method 2" views large changes in earnings as a *temporary* shift in an individual's wage profile[24]. Thus individuals' wage *history*, rather than just the most recent observation, is now informative for estimation of future wages. A hypothetical example is shown in Figure 5.

FIGURE 5 about here

To implement this method, I use a fixed effects regression model following the methodology of Carneiro and Heckman (2003). Appendix 2 describes this prediction method in detail, including model specifications and robustness checks. I also show in Appendix 2 that this produces *average* age 30 wage estimates that are very similar to those from "Method 1" (though wage estimates from method 2 suffer from less variability)[25]. The similarity of average wage predictions across methods is due to the "shocks" that are incorporated in method 1 being both positive and negative (hence cancelling each other out on average).

In Table 7, I compare my predictions from these two methods to similar information recorded for 30 year olds in an external data source (the 2003-2005 CPS March Annual Supplements[26]). In general, average predicted age 30 wages appear similar to those in the external CPS data. I predict average wages to be \$29,600 in the NELS, while in the CPS the equivalent figure is \$28,900. Likewise, my predictions of average wages seem to be reasonably close to the CPS data for a number of sub-groups (e.g. those who are white or holding a bachelors degree), though there are instances where this is not the case (e.g. those with less than high school education and Hispanics)[27]. It is worth noting, however, that both

[24] I am using the term "temporary" in a slightly different manner here compared to page 19. Specifically, for the illustrative individual in Figure 3 I do not assume that their wage growth reverts to their age 23-25 trajectory (and thus that his wage at age 26 contains no useful information in predicting age 30 wages at all). Rather I allow the age 26 wage to have some permanent impact on my prediction of his age 30 wage, but for it to be tempered by what they were earning between ages 23 and 25.

[25] This is because outlying observations are moderated in Method 2 by the influence of previous wages (it is a time mean). This does not occur in Method 1, where it is only the most recent observation that is used for prediction. Hence if there is a large shock to the most recent observation, this gets carried forward to the future prediction in Method 1, as opposed to being averaged out in Method 2.

[26] The exact wording to collect income and wage data in the CPS is comparable to that used in the NELS.

[27] I predict those with less than high school education to earn around \$21,000 while the CPS figure stands at just over \$15,000. My definition of "less than high school education" is those who made it into the final year of high school but did not graduate. On the other hand the CPS represents the whole US population, and defines less than high school education as everyone who did not graduate from high school, *including* those who dropped

these groups only form a small part of the overall sample. The general message is that my prediction methods seem to generate a reasonable estimate of *average* age 30 wages. Further evidence of this can be found in Appendix 2. However, in Figure 6, I show that my two wage predictions for the illustrative NELS respondent are $15,000 (30%) apart. Anywhere between the two predictions, or even a figure outside of this range, could be possible. Thus a comparison of expected and actual wages at the *individual* level does not seem a sensible approach with the NELS data. On the other hand, when dealing with group averages, over-estimates of wage growth for some individuals will be compensated by underestimates for others.

TABLE 7 & FIGURE 6 about here

Unearned Income

At age 26, respondents' were asked about their non-wage income at age 25, with 74% of individuals reporting no unearned income[28]. Unearned income may make up a more significant proportion of total income at age 30 than at age 26. To investigate this, I compare mean wages to the mean total income for 30 year old men in the 2003-2005 CPS March Annual Supplement[29]. Mean total income for this group is only $500 higher than mean wages. This suggests that "other" sources of income make up only a small fraction (roughly 2%) of 30 year old men's total income (on average). I also investigate the extent of unearned income reported in another American data source (the NLSY 1979), again finding that it has very little impact on the average individual (the median unearned income is zero)[30]. Therefore, to incorporate unearned income into my predictions, I simply use the value recorded at age 25 in the NELS. Given the minor contribution this makes to individuals total income, this should not introduce substantial bias at the group or population level (the same,

---

out *before* their senior year. This is probably the reason why, in the NELS comparing to the CPS, my predicted wage is higher and there are a smaller proportion of respondents with below high school education. In a similar manner, I predict average wages for Hispanics to be around $27,000, while the CPS figure is closer to $20,000.

[28] The exact wording of this question can be found in Appendix 2, and asks respondents to include income from savings, stocks and bonds along with any child support, family or disability payments.

[29] Several questions about other (unearned) sources of income were asked in the CPS. This includes how much they received from benefits, welfare, assistance, dividends and interest. Hence the definition of "other income sources" seems largely comparable with that applied in the NELS (though an obvious difference is that the NELS asks for this information in a single question, compared to several component parts in the CPS). The data I use is drawn from the CPS "Table Creator", available from http://www.census.gov/hhes/www/cpstc/cps_table_creator.html I produce two values, one looking at men's average wages, the other their total income. I assume that the difference between these figures (average wages and average income) equals total income from unearned sources.

[30] Infact 62% of men report having no unearned income.

however, is unlikely to be true if one were to try and make inferences at the individual level).

Summary

I have presented two methods to predict age 30 wages, both of which are comparable with external estimates from population level data. Moreover, even though age 30 unearned income is difficult to predict, this makes up only a small proportion of total average income at the group or population level. I am therefore confident that the substantive inferences in section 5 regarding population and group level averages are robust to the data issues discussed throughout this section. However, inferences made at the individual level are likely to suffer from what may be quite severe biases. Thus I choose not to conduct such analysis in this paper. For a more detailed discussion of these issues, I encourage the reader to turn to Appendix 2, where I present a full description and justification of the methods used.

5. Results

In this section, I will compare students' expectations to my prediction of their income at age 30. Throughout this discussion, I will focus on the results using prediction "Method 2" (the fixed effect extrapolation model) from the previous section. Results using "Method 1" are generally consistent to those presented, with a discussion in Appendix 3.

Before investigating the relative accuracy of students' and workers' expectations, Figure 7 presents the distributions for expected and predicted age 30 income across all individuals (i.e. both students and workers).

Expectations (dotted lines) are clearly to the right of the predicted age 30 income distribution (solid lines). Very few 20 year olds expect to earn less than $20,000 at age 30, though I predict that almost a quarter do. Conversely, there is quite a significant minority (3%) expecting to earn $100,000 or more, though in reality very few (1%) reach this milestone. Indeed, the median predicted income is $26,695 compared to expectations of $40,000, an average overestimation of around 50%[31].

[31] Note that here I am discussing the median. In Table 7, where I compared predicted age 30 wages to data from the CPS, I am discussing the mean.

FIGURE 7 about here

I check the robustness of this result in Table 8, which illustrates the proportion of adults expecting to be in each occupation by age 30, and the actual proportion in each by age 26[32]. The last column gives the median wage for workers of all ages in each occupational group in 2004, drawn from CPS data, to give an idea of the financial status of each occupation (note Table 8 is ordered by this column).

TABLE 8 about here

Although one can only make quite a crude comparison, as the data relate to expectations and realisations at different ages, it nevertheless illustrates that young adults also seem to be overly ambitious in their occupational expectations[33]. There are fewer individuals in the highest paying occupations (engineers, arts, doctors) and more in the less well paid (sales, services and clerical roles) than expected. Moreover, note that in the column labelled "Difference between expected and actual", negative figures tend to sit near the top of the table and positive numbers at the bottom. This also suggests that young adults occupy lower paying jobs than they previously expected.

To summarise Table 8, I derive an "expected" and "predicted" income from this occupational data. Specifically, I use the reported proportions in each occupation as weights (i.e. column 2 as weights for expected income, column 3 for actual income), which I multiply by the occupation specific CPS wage (column 4). Using this method, I find that young adults expect an income of $29,683, but I predict them to actually obtain $24,538. Hence they overestimate their future income by 20%. Though this figure is significantly below the 50% found above, one should remember that this method captures just one aspect of the underlying issue. Even

[32] Of course, some young adults are still in education at age 26, who are likely to be working professionals by age 30. However, this group is only relatively small, and are contained within the 4.5% described as "not working/studying/homemaker". Even if I assigned this group to a professional category, I would still find large overestimation in the results.

[33] Of course, there is an issue that I observe individuals at 26 rather than 30, and they could change occupation over the unobserved period. However, if respondents were asked what occupation they expected to have at age 26, would one really expect them to give a substantially different answer? I believe that expected occupation at 26 and expected occupation at age 30 would be highly correlated, and for this to be a reasonable proxy.

though a young man may be able to predict his occupation, he may overestimate the general pay that is received in that profession, or expect to be further up the career ladder than he actually achieves (e.g. expecting to become an army Sergeant by age 30, but only ending up a Private).

In general, this preliminary analysis strongly suggests that young men overestimate their future labour market success. They expect an average income of $40,000, but in reality I predict their annual earnings to be less than $30,000. Moreover, many young men expect professional work that they do not go on to achieve.

The accuracy of students and workers

Figure 8 presents results, analogous to the above, for just those sample members who were still in education at age 20. Clearly, the results are very similar to those presented above. The median predicted income (using Method 2) is $30,187 compared to expectations of $45,000; students overestimate their future income by around 50%. Likewise, Table 9 illustrates how the career expectations of young men still in education match with their eventual occupational attainment by age 26.

FIGURE 8 & TABLE 9 about here

As in Table 8, many students expect to work in professional careers, but end up in less prestigious jobs. For instance, whereas 8.5% expect to become artists or entertainers, only 2.5% work in these occupations by age 26. Likewise, only 6.2% are engineers at age 26, though around 11% thought they would be working in this profession. Again, when using the data in columns 2 and 3 to weight column 4 (as on page 24), I find students expect to earn, on average, $33,465 but their actual average income is $27,097; overestimation of around 25%.

There seems sufficient evidence to conclude that students overestimate their future success in the labour market. I find they overestimate their income at age 30 by, on average, 50%. Likewise, I find that many graduates are working in service, clerical and sales roles that as 20 year old students they did not anticipate doing for a career. These findings complement results from Smith and Powell's (1990) study of two mid-western universities. They found students overestimated their salary at age 30 by around 40%. It seems that this general result holds across the wider US student population.

I now turn to the results for young adults who were already in the labour market when asked for their income expectations. In section 2, I argued that:

(a) Workers may make more accurate predictions of future income than their student peers

and

(b) That their expectations may, on average, be realistic.

Figure 9 shows little support for either of these hypotheses. Those who were working at age 20 expected an age 30 mean income of $40,000. In reality, I predict their mean income to be $24,789. Workers are overestimating their wage, on average, by around 60%. This is similar to the overestimation made by students, where I found a figure of 50%. These results are supported by my investigation of workers' occupational expectations in Table 10.

FIGURE 9 & TABLE 10 about here

Whereas 4.7% of workers expected to become an engineer, less than 1% were working in this occupation at age 26. On the other hand, around 1 in 20 thought they would be working as a labourer by age 30. Yet around an eighth held this job at age 26. Calculating weighted average wages from this occupational data (see page 24), I find workers expect a wage of $29,263 but end up receiving $25,907; a difference of 15%. Once again, this may appear to be small when compared to the 60% overestimation in wages. But I remind the reader that this method captures just one aspect of the underlying issue (as discussed previously on page 24).

Overall, there is little evidence that workers hold realistic expectations. In fact, on average they are just as unrealistic as their student peers. Both groups tend to overestimate their future income and occupation; many believe they will receive the financial rewards on offer in professional careers, but will ultimately not be able to obtain this goal.

Comparing workers to different groups of students

The above analysis has treated those in higher education at age 20 as a homogenous group. In reality, students differ in all manner of characteristics, including the subject they study, prior academic achievement, whether they also hold a job, and, looking into the future, whether they eventually graduate. I extend the above analysis by trying to answer three questions. Firstly, though workers may not hold more realistic expectations than the "average student", they may make better predictions than particular groups. Do workers make better predictions than say Art and Humanities students, for instance, whom I found to be the least realistic over short time horizons in the UK? Secondly, are factors such as race, ability and social class influencing both enrolment in higher education and the accuracy of expectations? If so, is it these factors that are driving my results? Finally, can I provide any further evidence that experience in the labour market is unrelated to accuracy of expectations, as my findings so far suggest, by considering differences between students with and without a job?

I investigate these questions by estimating an Ordinary Least Squares regression model[34]. I specify the dependent variable as the natural logarithm of the expected income divided by the predicted age 30 income:

$$\log\left[\frac{Y_{ij}^{exp}}{Y_{ij}^{act}}\right]$$

Webbink and Hartog (2004) use a similar specification in their analysis of Dutch students' wage expectations. This specification is assumed to satisfy the condition that errors are normally distributed with constant variance[35]. It also allows a distinction between respondents who over and under estimate their future income, unlike the specification preferred by Betts (1996) and Wolter (2000). Later in this section, I also present quantile regression estimates as an alternative to test the robustness of my results[36].

[34] One might suggest that this looks like an individual level analysis that I ruled out in the previous section as inappropriate. On the other hand, another way of looking at this is that I am analysing conditional means (and thus that I am infact undertaking further analysis at the group level).
[35] Analysis of the OLS residuals was carried out thoroughly after estimation of each regression model. There was little evidence that the normality and constant variance assumptions were violated.
[36] I choose to present the OLS results as I can easily take into account the complex survey design used in NELS. This is much trickier when using alternatives such as quantile regression.

In the first specification, I compare working 20 year olds to students defined by the subject that they study[37]. In other words, workers enter the regression model as the reference group, with 14 dummy variables representing students in different disciplines. I then add in a term reflecting the respondents' cognitive ability in mathematics at age 18. Intelligent individuals are more likely to enter higher education. But they may also be particularly adept at processing the labour market information that they receive. Once I have controlled for students' superior ability, do I find that workers (who may hold a greater quantity of information) hold more realistic expectations?

Specification 3 controls for a series of other potentially confounding factors, including race and family background. Work from the sociological literature, for example Baird et al (2008), describes the importance of controlling for these characteristics when considering the accuracy of students' expectations. I also include an indicator for whether the respondent was a student who also held a part-time job while at university. Previously, I argued that workers will have more accurate expectations than students as they hold more labour market information. In a similar manner, one would expect students in paid employment to hold more labour market information than their university peers without a job. I use this analogy in the final specification as an additional test of whether labour market experience is related to accuracy of income expectations.

However, when making comparisons between students and workers, one should remember that some of those who enter higher education leave without obtaining a degree. In other words, even though these students were enrolled in a college programme at age 20, they may not have obtained a university level qualification by age 26. The OECD 1998 Education at a Glance report (OECD 1998) notes that the USA has a relatively low university completion rate (just over 60% of those who enter). This is reflected in the NELS data; around 30% of those who were students at age 20 had not obtained a degree by age 26[38]. I take this into account in my final specification by including a dummy variable that indicates whether the individual became a "college drop-out" (i.e. enrolled in university at age 20, but had not obtained a degree by age 26). Interpretation of the subject dummies will therefore change

---

[37] Under both prediction methods, students in all subjects are assumed to have the same average annual real wage growth rate between 26 and 30. I also experimented with a prediction model that allowed wage growth to vary between graduates from different subjects. Results were largely the same to those presented.

[38] Morgan (2005) finds a similar proportion when he uses a different sample selection of the NELS data.

between the first three specifications and the fourth. In particular, the final specification will indicate the accuracy of students' expectations compared to workers, *conditional* on successful degree completion by age 26.

Formally, the final specification of the model is:

$$\log\left[\frac{Y_{ij}^{exp}}{Y_{ij}^{act}}\right] = \alpha + \beta_0 R_i + \beta_1 F_i + \beta_2 S_i + \beta_3 T_i + \beta_4 D_i + \beta_5 W_i + \varepsilon_{ij}$$

With:

Y_{ij}^{exp} = Expected income at age 30

Y_{ij}^{act} = Predicted income at age 30, using Method 2

R = Race

F = Parental income when respondent was 18 years old

S = Subject of study, working or unemployed at age 20

T = Measure of individual ability at age 18

D = Whether the respondent was a university student at age 20, but had not obtained a degree by age 26

W = An indicator of whether the individual was a student who also held a job at age 20

ξ_{ij} = Error term. Individuals were initially sampled by school clusters at age 14, which is accounted for by adjusting the standard errors.

i = Individual i

j = School j, that the individual was initially sampled from at age 14. All standard errors have been adjusted to take into account the complex sampling design (clustering of children within schools)[39].

Results are presented in Table 11. Model 1 enters just the indictor of whether the respondent was working or a student, defined by the subject they were studying, at age 20. The results show workers sit somewhere in the middle of this ranking; they make better predictions than some students, but worse than others. Engineering, Physical Sciences, Maths, Education and Agriculture students are all more realistic on average than workers at the 5% level. Similarly,

[39] I also experimented with a fixed effects regression model, including a dummy variable for each school that children were initially sampled from. Results were largely unchanged from those presented.

Art, Law, Journalism and Biological Science students are all less realistic than workers at the 5% level. Figure 10 shows this in more detail, highlighting by how much each group overestimates their age 30 income (on average)[40].

TABLE 11 & FIGURE 10 about here

It appears that young adults who are studying vocational, financial and mathematically based courses hold reasonably realistic expectations. For instance, those studying Computer Science and Maths overestimate their age 30 income by a (comparatively) small 20%. Students in Agriculture are even more realistic, their expectations are only just statistically different to their predicted realisations. Yet those studying artistic and writing based courses, with the exception of Language students, expect almost double what I predict them to earn.

This ranking of subjects is very similar to my results for UK students' starting salary expectations, presented in Jerrim (2008). Thus across countries, over a long and short time horizon, it appears that vocational and mathematical students are the most realistic, while those in more creative subjects are the least. It is interesting to consider this result in light of studies that have investigated actual wage differentials by college major. Specifically, there are some quite striking consistencies with Black et al (2003). The authors of this study note that perspective students are provided with 'little concrete information' about the labour market success of graduates from the array of different US subject majors. Hence they illustrate the wages of graduates in around 40 disciplines, relative to the earnings of those who have left university with an Economics degree. They find that Engineers receive the highest wages, just as I find them to be the most realistic. Likewise, I find Art students are amongst the least realistic, while Black et al show that this group earn the lowest wages. Some other patterns seem to hold too; for instance when drawing comparisons between Biologists (lower earners and less realistic) and Physicists, Mathematicians and Chemists (higher earners, more realistic).

[40] To calculate how much a person with given characteristics overestimates by, one must sum the relevant coefficients from Table 11, and then take the exponent of this value. For example, take a journalism student. I want to know how much they overestimate their future income by using specification 1. Firstly, I sum the relevant coefficients (0.41+ 0.18) to get 0.59. I then take the exponential of this value exp(0.59), to get a value of 1.80. If I then subtract 1 from this value (1 is where expectations equals realisation) and multiply by 100 (to get the value into percentage form), I arrive at the average % overestimation (80% shown in Figure 10).

In the second model I control for respondents' cognitive ability in maths on a test taken at age 18. Earlier, I suggested that those of higher ability maybe more realistic. This seems to be consistent with the data. A worker of average ability overestimates his future wage by 46%. However if their maths test score was two standard deviations above the mean, I predict they overestimate their future income by only 25%. Notice that the subject dummy coefficients have slightly decreased, for instance from -0.22 to -0.16 for Engineering students and -0.09 to -0.02 for Economics. Hence, although it seems that cognitive maths ability is related to accuracy of young adults' income expectations, there is little evidence to suggest this is why I find that some groups of students to make better predictions than workers.

In model 3, I add additional controls for Race and Family background. As stated by Baird et al (2008), race influences the accuracy of expectations. Blacks, Hispanics and Latinos all make worse predictions than Whites. On the other hand family background, measured by parental income quintile, is statistically insignificant. Notice, however, that none of the substantial results from model 2 change. Differences in family backgrounds and ethnicity do not explain why I find no difference between students and workers. I also include a variable that indicates whether the respondent was a student who also held some sort of formal employment at age 20[41]. If work experience provides young adults with valuable labour market information, which they process rationally, one would expect this group to make better predictions than their student peers who do not have a job. Again, this does not seem to be the case. The coefficient is small and statistically insignificant, with this result holding across several specifications and unconditional estimates not presented. This supports my finding that labour market experience is unrelated to the accuracy of young adults' long term income expectations. Not only do young adults working full-time make no better predictions than students, but those enrolled in higher education seem to receive little information about their long-term prospects in the labour market from holding a part-time job.

The final specification enters a dummy variable for individuals who were students at age 20, but had not obtained a degree by age 26[42]. In other words, this group dropped out of university *after* reporting their income expectations. Firstly notice the large (0.31) and highly significant ($t \approx 6$) coefficient. These students make particularly poor predictions of their

[41] Around half of the students surveyed in the NELS fell into this category.
[42] This group makes up 17% of the 4,434 NELS sample members analysed in this paper.

future income, and are in fact significantly less realistic than their peers who entered the labour market straight from school[43]. For instance, workers overestimate their age 30 income by around 45%[44], compared to 130% for a 20 year old Art or Journalism student who failed to complete their degree. This result can be interpreted in two ways. One possibility is that these students stated their income expectation on the assumption they would obtain a certain level of human capital and a valuable labour market signal. However, they did not go on to actually receive the outcomes they initially anticipated from their human capital investments, thus causing their apriori expectations to be incorrect. Alternatively, these individuals could have dropped out of university *because* of their overly ambitious expectations. For instance, they may have gone to university thinking they would earn a high wage (i.e. their high expectation observed at age 20). But through their later experiences, they may have revised down their expectations substantially (i.e. If I were to observe expectations at age 21 say, they would be much lower). On the basis of this revision, they have decided that the benefits from obtaining a degree are not worth their continued investment, and hence leave university before the end of their course. Hence this variable is potentially endogenous; it could be these students unrealistically high expectations that is driving their decision to leave university, rather than their expectations being unmet because they drop out.

Figure 11 illustrates how including this variable leads to a large change in the other parameter estimates. There is a particularly large effect on the subject dummy coefficients. Recall that these now compare the accuracy of students' expectations to workers, conditional on whether they complete university by age 26. The light grey bars in Figure 11 are the estimated overestimation of age 30 income for each group, calculated using specification 4. These results add further weight to my finding that students are no less realistic than workers. Consider, for instance, Business students. In specification 1-3, they were statistically indistinguishable from workers, overestimating their age 30 income by around 45%. Now I find they are substantially *more* realistic, conditional on them having graduated by age 26, overestimating their age 30 income by a comparatively small 30%. On the other hand, those who dropped out of Business school overestimate their future income by around 75%. Similarly, the once statistically significant difference between Art, Law, Health and

[43] For the prediction of age 30 income in section 4, I have treated this group the same as those with a high school qualification who never been to university at all. One may argue that college drop outs may have some wage premium over this group. All the results from specification 2 onwards still hold even if I make different assumptions (e.g. that this group have the same wage growth rate as those with an associate degree).
[44] Note this is lower than before. Recall previously I focused on the median. Now I am using OLS regression, the measure of central tendency used is the mean.

Journalism students and workers has now disappeared. There is even less evidence that workers are more realistic than students, and in fact quite the opposite may even be true.

It is also interesting to note that, for some groups of students, the difference between expectations and later realisations is now statistically insignificant. This includes Agriculture, Computer Science, Maths, Engineering and Physical Science graduates, who overestimate their age 30 income by, on average, less than 10%. Hence, conditional on successful completion of an undergraduate degree, I find that some groups of students are actually quite realistic, even over a relatively long time horizon.

FIGURE 11 about here

In Table 12 I present various robustness tests that refer to the third specification of the regression model described above[45]. Specifically, model A refers to when I do not extrapolate the actual wage data, and simply compare age 26 income to expectations at age 30. Model B presents results when using "method 1" to predict age 30 wages, as described in section 4. Model C adjusts for the item non-response described in Tables 1 and 4, via the application of response weights I have created from a logistic analysis of missing data, while model D refers to quantile (median) regression estimates.

TABLE 12 about here

The results generally support those presented in Table 11. Notice that Blacks and Hispanics always make worse estimates than Whites, while family income is never statistically significant. Likewise, higher ability is always associated with more realistic expectations. The estimated subject dummy coefficients are also similar to before; Art, Biological Sciences and Journalism students make relatively poor predictions compared to both workers and their university peers who are studying Agriculture, Engineering or Physical Sciences. Likewise, in analysis not presented, I found the "college dropout" variable to be positive, strong and

[45] I have chosen this specification as it controls for the largest number of pre-determined factors, without including the potentially endogenous "college dropout" variable. In analysis not presented, I ran each of these robustness tests for all specifications and found largely consistent results.

highly significant, while its inclusion again caused a relatively steep decline in the other model coefficients.

To further test the robustness of my results, I estimate a binary logistic regression model of whether the respondent, at age 26, was working in the occupation he expected to be in at age 20. As argued previously, those who are the most realistic about their future occupation should also be the most realistic about their future income. Consequently, one would expect to see results from analysis of occupational data to be consistent with the above results regarding income. For instance, I should find that maths students make better predictions of their future occupation than those from creative subjects. Likewise, individuals who drop out of university should only rarely enter the occupation they expected at age 20.

Specifically, I estimate a logistic regression using the binary indicator O as the response. This variable is assigned the value 1 if the respondent, at age 26, was working in the occupation he expected to be in at age 20. I enter the same covariates as in the model described on page 30 and Table 11. Formally, this model is specified:

$$\log\left(\frac{\Pi(O_i)}{1-\Pi(O_i)}\right)=\alpha+\beta_0\mathrm{R_i}+\beta_1\mathrm{F_i}+\beta_2\mathrm{S_i}+\beta_3\mathrm{A_i}+\beta_4\mathrm{D_i}+\beta_5\mathrm{W_i}+\varepsilon_{\mathrm{ij}}$$

Where:

$\Pi(O_i)$ = Probability of respondent i entering the occupation they expected at age 20 (by age 26)

R = Race

F = Parental income when respondent was 18 years old

S = Subject of study, working or unemployed at age 20

A = Measure of individual ability at age 18

D = Whether the respondent was a university student at age 20, but had not obtained a degree by age 26

W = An indicator of whether the individual was a student who also held a job at age 20

ξ_{ij} = Error term. Individuals were initially sampled by school clusters at age 14, which is accounted for by adjusting the standard errors.

i = Individual i

j = School j, that the individual was initially sampled from at age 14

Results are presented as odds ratios in Table 13. A coefficient greater than 1 indicates a higher probability of the respondents' occupational expectations being correct. There is reasonable agreement between these results and those presented in Table 11. For instance, notice that mathematical and vocational students tend to make better predictions of their future occupation than workers, just as Table 11 showed they made better predictions of their income. However, it is also worth pointing out that, in some specifications, the coefficients only begin to approach traditional levels of statistically significance (though, qualitatively, they have the same sign). It is also interesting to note how the inclusion of the "college drop out" variable in specification 2 has the same effect in Tables 11, 12 and 13 (the subject coefficients all tend to increase, while the expectations of university drop-outs are particularly unlikely to become true).

TABLE 13 about here

Qualitatively, other results from Table 11 also hold. Lower ability respondents are less likely to enter the occupation they expected at age 20, though this does not quite reach statistical significance, even at the 10% level. Likewise, Black respondents are less likely to enter their chosen occupation than Whites, but again this is not statistically significant. One interesting difference is that the coefficient on students holding a part-time job is now statistically significant at the 5% level. However, it is the opposite sign to what one would expect; those with a part-time job are *less* likely to enter the job they expected.

Nevertheless, the directions of the parameter estimates are generally consistent with the results presented in Table 11. It seems that (qualitatively) these results support my substantial conclusions from the income expectations data. This gives me further confidence that my results are not being driven by my prediction methods for age 30 income or assumptions I make about the income expectation data.

3.6. Discussion and conclusion

The small US and European literature on young adults' expectations has typically focused on how well university students can predict their first salary upon graduation. Though they sometimes deal with longer time horizons, results are normally shown to hold for only a very specific proportion of the US student population. Furthermore, existing work rarely compares the accuracy of different groups. There is little or no comparison of students versus workers, those studying for a mathematical degree versus a more creative subject, or those who successfully graduate versus those who do not. I try to resolve these issues by using rich, longitudinal data that has been drawn from across the US population of high school seniors. Hence I not only make a better attempt at representing the expectations of the US student population, but also tackle a set of new and interesting hypotheses that have not been previously considered in the literature.

However, one should not ignore the difficulties I have encountered with the NELS data. Missing data, particularly the fact that age 30 income is not directly observed, is a notable problem. A second issue is whether the questions asked are accurately capturing young adults' expectations (what they realistically believe will happen) rather than their aspirations (their hopes and dreams). The ordering and wording of the questions (given in section 3) should have guided respondents towards making a realistic assessment of their future income. Yet certain groups may have interpreted this question quite literally (e.g. Maths students report their expectations) while others have not (e.g. Art and Journalism students state their aspirations). This issue is not specific to this paper, but rather the more general practice of collecting expectation data in economic research. As such, this seems an area that is ripe for future work.

Noting these caveats, my results suggest that, on average, students at age 20 have unrealistic expectations of their income at age 30. Yet this broad result needs qualification. Certain groups of students, conditional on successful completion of their degree, are actually quite realistic. For instance, Maths, Education and Engineering students overestimate their age 30 income by less than 10%. One may wish to view these results in light of Black et al (2003), who note that US students are provided with 'little concrete information' about the success of graduates in different subjects. Their paper tries to provide this information, illustrating the wages of graduates in around 40 disciplines. They find that Engineers receive the highest wages, just as I find them to be the most realistic. Likewise, Art students are amongst the lowest earners and least realistic. Even some quite specific patterns seem to hold. Biologists earn less than Physicists, Mathematicians and Chemists, though they do not seem to realise this when undertaking their human capital investments. Consequently, my results suggest there is certainly a need to provide prospective students with the type of information presented by Black and his co-authors.

On the other hand, there is substantial evidence that those young adults who are working at age 20 make quite poor predictions of their future income; overestimation is, on average, 50%. It is also interesting to consider again why workers seem to make no better predictions than students. As stated at the start of this paper, it maybe that young workers are not focused on a particular career and hence suffer from a lack of direction in the labour market. Alternatively, it might be that young workers are myopic and choose to collect information from those who are closer to them in terms of age and the next rungs on the career ladder. Another possibility is that workers have both "accurate" and "inaccurate" sources of labour market information that they struggle to distinguish between. For instance, a manager may be keen to retain a particular staff member who is considering employment elsewhere. Thus the manager may overstate the chances of pay and progression within the firm. If the worker can not tell that this is "bad" information, it may lead him to raise his future income expectations. Indeed, in situations where workers only receive relatively poor quality information, one would expect them to be no more (and possibly even less) realistic than their student peers.

Finally, some young adults may not realise the value of the information that they hold, or how it applies to them and their future; they may discard (or give less weight) to some important information as they see it as irrelevant. For example, a young worker may know what a 30 year old employee in his organization is paid. But he (perhaps unrealistically) views his current job as a stop-gap solution, and believes he will have entered an entirely different industry in a few years time. He therefore does not fully incorporate the information he holds on the wages of 30 year olds into his income expectations. Indeed, this interpretation seems to be consistent with the findings of Smith and Powell (1990). They find that students can accurately estimate average graduate wages, but expect their own salaries to be a lot higher. Hence, although they are well informed about average wages (i.e. hold relatively good information), they do not necessarily incorporate this into their expectations (i.e. make good predictions of their own future salary).

Linking these points back to my opening paragraphs, simplistic assumptions that young adults hold a combination of "full-information" and "rational expectations" may be based on a rocky foundation. It seems that young adults may be missing some important labour market information, making further research in the spirit of Black et al (2003) ever the more important. Yet economists must also develop a better understanding of how young adults use the labour market information that they hold. In many ways, it is difficult to believe they will give it the appropriate weight when making schooling decisions as is often assumed in a rational expectations framework; indeed, as the quote from Becker suggested at the start of this paper, young adults probably do not realize their own limitations and tend to over-estimate their chances of good fortune[46]. Thus understanding exactly what information young adults hold, how they use it, and the effect this has on their schooling decisions should become an important area of future economic research.

[46] See Chevalier et al (2009) for some empirical evidence on this topic.

References:

Baird, C. Burge, S. & Reynolds, J. (2008) “Absurdly Ambitious? Teenagers' Expectations for the Future and Realities of Social Structure”, Sociology Compass, volume 2, p 944-962

Becker, G. (1993) “Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education”, Third Edition, University of Chicago Press

Berger, M. (1988) “Predicted Future Earnings and Choice of College Major”, Industrial and Labor Relations Review, volume 41, p 418-429

Betts, J. (1996) “What Do Students Know About Wages? Evidence from a Survey of Undergraduates”, Journal of Human Resources, volume 31, p 27-56

Black, D. Sanders, S. & Taylor, L. (2003) “The Economic Reward for Studying Economics”, Economic Inquiry, volume 41, p 365-377

Blau, F. & Ferber, M. (1991) “Career Plans and Expectations of Young Women and Men: The Earnings Gap and Labor Force Participation” Journal of Human Resources, volume 26, p 581-607

Borjas, G. (1999) “The Economic Analysis of Immigration”, Handbook of Labor Economics, p 1697-1760

Boudarbat, B. (2004) “Earnings and Community College Field of Study Choice in Canada”, IZA Discussion Papers 1156

Bound, J. Brown, C. & Mathiowetz, N. (2001) “Measurement Error in Survey Data”, in Heckman, J. & Leamer , E. (ed.) Handbook of Econometrics, edition 1, volume 5, chapter 59, p 3705-3843

Brunello, G. Lucifora, C. & Winter-Ebmer, R. (2001) “The Wage Expectations of European College Students”, CEPR Discussion Papers 2817

Brunello, G. Lucifora C. & Winter-Ebmer, R. (2004) “The Wage Expectations of European Business and Economics Students”, Journal of Human Resources, volume 39, p 1116-1142

Carneiro, P. & Heckman J. (2003) “Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College”, NBER Working Paper 9546

Carstensen, L. & Woltman, H. (1979) “Comparing Earning Data from the CPS and Employer’s Records”, Proceedings of the Social Statistics Section, American Statistical Association, p 168-173

Carvajal, M. Bendana, D. Bozorgmanesh, A. Castillo, M. Pourmasiha K. Rao, P. & Torres, J. (2000) “Inter-Gender Differentials Between College Students’ Earnings Expectations and the Experience of Recent Graduates”, Economics of Education Review, volume 19, p 229-243

Chevalier, A. Gibbons, S. Thorpe, A. Snell, M. & Hoskins, S. (2009) “Students’ Academic Self-Perception”, Economics of Education Review, volume 28, p 716-727

Curtin, T. Ingels, S. Wu,S. & Heuer, R. (2002) “National Education Longitudinal Study of 1988, Base-Year to Fourth Follow-up Data File User's Manual”, National Center for Education Statistics

Dominitz, J. & Manski, C. (1994) “Eliciting Student Expectations of the Returns to Schooling”, NBER Working Paper 4936

Dominitz, J. & Manski, C. (1996) “Eliciting Student Expectations of the Returns to Schooling”, Journal of Human Resources, volume 31, p 1-26

Dominitz, J. (1998) “Earnings Expectations, Revisions, And Realizations”, The Review of Economics and Statistics, volume 80, p 374-388

Jerrim, J. (2008) “Wage Expectations of UK Students: How Do They Vary and Are They Realistic?”, s3ri working paper series A08/08

Minnicozi, A. (2003) “Estimation of Sons’ Intergenerational Earnings Mobility in the Presence of Censoring”, Journal of Applied Econometrics, volume 18, p 291-314

Morgan, S. (2005) “On the Edge of Commitment: Educational Attainment and Race in the United States”, Stanford University Press

Murphy, K. & Welch, M. (1990) “Empirical Age-Earnings Profiles”, Journal of Labor Economics, volume 8, p 202-229

Organisation for Economic Co-operation and Development (1998) “Education At a Glance: OECD Indicators 1998”, OECD Paris

Rouse, C. (2004) “Low-Income Students and College Attendance: An Exploration of Income Expectations”, Social Science Quarterly, volume 85, p 1299-1317

Rubinstein, Y. & Weiss, Y. (2006) “Post Schooling Wage Growth: Investment, Search and Learning”, In Hanushek, E. & Welch,F. (eds.), Handbook of the Economics of Education, volume 1, p 1-67

Smith, H. & Powell, B. (1990) “Great Expectations: Variations in Income Expectations Among College Seniors Sociology of Education”, volume 63, p 194-207

Webbink, D. & Hartog, J. (2004) “Can Students Predict their Starting Salary? Yes!”, Economics of Education Review, volume 23, p103-113

Willis, R. & Rosen, S. (1979) "Education and Self-Selection", Journal of Political Economy, volume 87, p S7-36

Wolter , S. (2000) “Wage Expectations: A Comparison of Swiss and US Students”, Kyklos, volume 52, p 51-69

Table 1. Logistic regression of item non-response to question on income expectation

	Specification 1		Specification 2		Specification 3	
	Odds ratio	**Standard error**	**Odds ratio**	**Standard error**	**Odds ratio**	**Standard error**
Socio-economic status index (Ref: Lowest quartile)						
Second quartile	0.85	0.22	0.85	0.21	0.86	0.20
Third quartile	0.58*	0.16	0.56*	0.16	0.59	0.16
Top quartile	0.62	0.16	0.59*	0.16	0.62	0.16
Family income parents reported when respondent was age 18 (Ref: Bottom quartile)						
Second quartile	0.47*	0.11	0.48*	0.11	0.48*	0.11
Third quartile	0.40*	0.10	0.41*	0.11	0.40*	0.10
Top quartile	0.24*	0.07	0.22*	0.07	0.22*	0.07
Race (Ref: White)						
American Indian or Alaska Native	4.52*	2.69	4.13*	2.17	3.16*	1.44
Asian or Pacific Islander	1.44	0.67	1.16	0.56	1.39	0.59
Black, not Hispanic	0.80	0.25	0.75	0.23	0.68	0.20
Hispanic or Latino	1.27	0.26	1.24	0.26	1.24	0.28
More than one race	0.52	0.18	0.54	0.18	0.50	0.18
Reported health problems at 20 (Ref: Yes)						
No	1.00	0.28	1.17	0.37	1.45	0.46
Labour force status at 20 (Ref: Student who does not have a job)						
Student who also has a job	0.84	0.20	0.77	0.19	0.83	0.20
Working only	1.93*	0.46	1.76*	0.41	1.76*	0.42
Not student or working	2.64*	0.73	2.28*	0.65	2.20*	0.65
Housing tenure at 26 (Ref: Homeowner)						
Rent from someone, not a relative	-	-	1.13	0.23	1.14	0.23
Rent from a relative	-	-	1.03	0.31	0.97	0.29
Live in residence without paying rent	-	-	1.77*	0.44	1.50	0.39
Wage at age 26	-	-	-	-	1.01	0.06
Maths ability	-	-	-	-	0.97	0.09
Drop out of university (Ref: No)						
Yes	-	-	-	-	1.07	0.28
Observations	**5,782**		**5,782**		**5,782**	

Notes:
1 The socio-economic status index is a continuous variable constructed by the survey organisers using data on the respondents' father's education level, mother's education level, father's occupation, mother's occupation, and family income. The higher this index score (or the higher the quartile), the more privileged the child's background.

2 An odds ratio greater than 1 means a greater chance of non-response relative to the reference.

3 "Wage at 26" is a continuous variable that records how much the respondent was earning from employment when they were aged 26. The estimated odds ratio shows how non-response increases with a $10,000 increase in the wage that was earned.

4 "Missing" dummy variables are included when the respondent has not provided information on a covariates. I do not present their results for brevity.

5 "Maths ability" is a continuous variable based upon respondents' scores in a test of their cognitive mathematical ability taken at age 18. The estimated coefficient in the table above shows how a one standard deviation *increase* in maths ability influences the propensity to not respond.

6 * Indicates statistical significance at the 5% level

Table 2. Timescale used to report salary

Timescale	% of observations
Hourly	10.4
Weekly	12.3
Twice monthly	3.4
Monthly	8.6
Annually	65.3
Observations	3,475

Notes:
1 I have restricted this data to those who were working full-time year round at age 26.

Table 3. Missing data on expected income and reported salary

	Observations remaining
All male respondents (Starting sample)	5,782
Individuals with missing expected income data dropped	5,149
Individuals with expected income below $6,000 dropped	5,110
Individuals with expected income over $250,000 dropped	5,039
Individuals with no full time wage observed between ages 23 and 26 dropped	4,434
Final sample	4,434

Notes:
1 Two item non response models are presented in Tables 1 and 4 that try to explain what factors are associated with missing data. Specifically, Table 1 investigates the drop in observations from 5,782 to 5,039 (missing or illogical expectations data). On the other hand, Table 4 looks at non-response to the actual salary data (i.e. the drop in observations from 5,039 to 4,434).

Table 4. Logistic regression of missing full-time wage history

	Specification 1		Specification 2		Specification 3	
	Odds ratio	**Standard error**	**Odds ratio**	**Standard error**	**Odds ratio**	**Standard error**
Socio-economic status index (Ref: Lowest quartile)						
Second quartile	1.06	0.31	1.05	0.31	1.06	0.36
Third quartile	0.93	0.24	0.88	0.22	0.96	0.28
Top quartile	1.09	0.34	1.02	0.33	0.84	0.30
Family income parents reported when respondent was age 18 (Ref: bottom quartile)						
Second quartile	0.77	0.16	0.79	0.16	0.80	0.19
Third quartile	0.59*	0.14	0.61*	0.14	0.63	0.17
Top quartile	0.32*	0.07	0.35*	0.07	0.37*	0.09
Race (Ref: white)						
American Indian or Alaska Native	0.55	0.57	0.72	0.74	0.91	0.88
Asian or Pacific Islander	2.86*	0.94	2.39*	0.73	1.59	0.44
Black, not Hispanic	1.03	0.33	1.07	0.33	0.82	0.22
Hispanic or Latino	0.85	0.24	0.86	0.25	0.79	0.24
More than one race	1.15	0.45	1.12	0.46	0.87	0.30
Reported health problems at 20 (Ref: Yes)						
No	0.37	0.16	0.41*	0.17	0.33*	0.22
Labour force status at 20 (Ref: Student who does not have a job)						
Student who also has a job	1.01	0.20	1.07	0.21	0.98	0.24
Working only	0.75	0.19	0.96	0.24	1.00	0.26
Not student or working	1.76	0.55	2.04*	0.63	1.84*	0.50
Housing tenure at 26 (Ref: Homeowner)						
Rent from someone, not a relative	-	-	2.25*	0.49	1.47	0.36
Rent from a relative	-	-	2.29*	0.87	1.52	0.70
Live in residence without paying rent	-	-	3.43*	0.82	1.82*	0.48
Expected income	-	-	1.01	0.01	1.01	0.01
Maths ability	-	-	1.27*	0.06	1.21*	0.07
Drop out of university (Ref: No)						
Yes	-	-	1.04	0.23	0.76	0.20
Working status at 26 (Ref: Working full-time)						
Work part time	-	-	-	-	4.57*	1.12
Study only	-	-	-	-	24.15*	6.27
Work full time & study	-	-	-	-	0.79	0.29
Work part time & study	-	-	-	-	15.88*	3.98
Neither work or study	-	-	-	-	18.93*	5.04
Observations	**5,039**		**5,039**		**5,039**	

Notes:

1 This table investigates the characteristics of the 605 young men who did not have a full time wage recorded at any point between the age 23 and 26

2 An odds ratio greater than 1 means a greater chance of non-response than the reference

3 See notes to Table 1 for details on the Socio-Economic Status Index and “Maths Ability” variables

4 “Expected Income” is how much a $10,000 increase in expected wage influences the chance of response.

5 * indicates statistical significance at the 5% level

Table 5. Summary statistics showing the NELS sample composition, before and after the exclusion of missing expectations and wage data

	Starting sample %	Final sample %
Labour force status at age 20		
Students who also have a job	26.6	26.0
Students who do not have a job	27.0	28.4
Working, not a student	35.0	35.8
Neither student or working	11.3	9.8
Highest qualification at age 26		
Less than high school	6.0	5.3
High school	55.6	56.2
Associates degree	7.1	6.9
Bachelors	28.2	28.6
MA/PhD	3.1	3.0
Race		
White	66.6	68.5
American Indian or Alaska Native	1.0	0.8
Asian or Pacific Islander	5.5	5.1
Black, not Hispanic	8.2	8.0
Hispanic or Latino	13.1	13.1
More than one race	3.0	2.6
Other	2.6	1.8
Family income student reported at age 18 ($1992)		
0-20000	18.5	18.1
20000-35000	19.8	19.8
35000-50000	17.7	18.4
50000-75000	16.0	16.5
75000+	12.3	12.4
Missing	15.7	14.9
University subject at 20 years old (If reported being a student)		
Agriculture	1.9	2.3
Accounting, Finance	6.1	6.2
Business Management	12.7	13.1
Journalism, Communication	3.3	3.6
Computer Science, Maths	4.8	5.4
Education	5.1	5.4
Engineering, Physical Sciences	16.9	17.4
Languages	1.8	1.7
Health	6.8	6.2
Law	4.2	3.9
Biological Science	7.4	6.6
Social Sciences, Humanities	9.1	9.1
Arts	5.0	4.6
Other	14.9	14.6
Working full-time At age 26		
Yes	74.0	84.0
No (e.g. unemployed, student, working part-time etc)	26.0	16.0
Observations	5,782	4,434

Notes:
1 “Starting sample” refers to all men in the age 26 sweep of the NELS. “Final Sample” refers to the sample I use in my analysis, once I have excluded missing data

Table 6. Average, annual (real) wage growth rates for young workers: Rubenstein and Weiss estimates

		% Average (real) wage growth rate per annum by education level				
Number of years experience in the labour force	**Data source**	**Below high school**	**High school**	**Some college**	**College graduates**	**MA/PhD**
0-10	CPS	2.4	3.2	3.3	3.6	2.9
	PSID	2.8	3.0	3.8	3.9	3.2
	NLSY	2.4	3.4	4.6	5.2	5.5
11-15	CPS	1.6	2.2	-	-	-
	PSID	1.9	2.0	-	-	-
	NLSY	1.3	2.3	-	-	-

Notes:
1 Source: Table 1, page 14 of Rubinstein and Weiss (2007) Post Schooling Wage Growth: Investment, Search and Learning. Handbook of the Economics of Education, Volume 1

Table 7. Predicted *mean* age 30 NELS wage compared to the *mean* age 30 Current Population Survey (CPS) wage

	% of observations in NELS	**Predicted wage method 1 ($000)**	**Predicted wage method 2 ($000)**	**% of observations in CPS**	**CPS wage ($000)**
Highest qualification at age 26					
Below high school	5.4	20.9	20.8	12.4	15.8
High school	56.4	27.0	25.5	47.8	24.7
Associates degree	7.0	29.9	30.5	8.2	28.1
Bachelors	28.0	37.7	37.8	24.6	38.1
Masters degree / PhD	2.8	42.4	43.6	6.8	44.5
Race					
White	69.8	31.4	31.3	60.6	31.5
American Indian	0.1	23.6	24.6	0.2	NA
Asian or Pacific Islander	5.1	38.5	36.4	6.2	35.4
Black (not Hispanic)	8.3	24.6	25.1	9.6	26.9
Hispanic	13.3	27.4	27.1	21.7	20.7
Other	4.6	27.1	26.6	1.7	NA
All respondents	100.0	30.4	29.6	100.0	28.9

Notes:
1 All observations in 1994 $
2 See notes to Appendix Table A8 for further details

Table 8. Proportion of 20 year olds expecting to enter each profession by age 30, and the proportion who have actually entered that profession by age 26

	% expecting to be in the occupation at age 30	% actually in each occupation at age 26	Difference between expected and actual (% points)	Average annual CPS wage for each occupation in 2004 (deflated to 1994 $000)
Homemaker, not working, studying	0.6	4.5	-3.9	0.0
Farmer	2.1	0.8	1.2	12.5
Labourer	2.0	9.3	-7.3	16.7
Service	1.1	3.1	-2.0	17.0
Skilled operative	3.1	7.7	-4.6	19.6
Clerical	1.5	6.2	-4.7	19.9
Craftsman	9.1	12.2	-3.1	20.6
Sales	2.3	6.3	-4.0	22.7
Protective services	7.7	3.7	4.0	24.9
Arts, Entertainment, Writing	6.8	1.7	5.1	29.3
Teacher	6.0	3.3	2.6	32.5
Professional Medicine (not Doctor)	5.5	1.2	4.3	35.0
Other Professional	11.8	6.0	5.9	35.6
Engineer	8.4	4.0	4.4	38.7
Computer technical	3.5	6.7	-3.1	39.2
Manager	11.4	12.7	-1.2	41.3
Legal	3.3	0.6	2.7	53.0
Doctor	3.4	1.1	2.3	63.7
Military	1.8	1.4	0.4	NA
Proprietor	8.5	7.3	1.2	NA
Observations	**4,218**	**4,368**		

Notes:
1 The difference column is the expectation % minus the actual %.

2 The number of observations differs due to missing data. In total, 4368 of the sample had an occupation recorded by age 26. Some of these individuals reported that they "did not know" what occupation they expected when asked at age 20 (hence a sample size of 4218)

3 The average CPS wage relates to the mean wage in each occupation for all workers above age 16 in 2004. This is the year the NELS sampled turned 30. I have included and ranked occupations by this information to give an objective measure of occupational status. Data is not available for military occupations and business owners Source: Table 39 http://www.bls.gov/cps/cpsa2004.pdf The weekly wages have been converted to annual equivalents and deflated to 1994 prices, using data from the US government social security office http://www.ssa.gov/OACT/COLA/awidevelop.html.

Table 9. Proportion of 20 year old *students* expecting to enter each profession by age 30, and the proportion who have actually entered that profession by age 26

	% expecting to be in the occupation at age 30	% actually in each occupation at age 26	Difference between expected and actual (% points)	Average CPS wage for each occupation in 2004 (deflated to 1994 $000)
Homemaker, not working, studying	0.6	4.8	-4.2	0.0
Farmer	2.0	0.7	1.2	12.5
Labourer	0.3	5.1	-4.9	16.7
Service	0.7	3.6	-2.8	17.0
Skilled operative	0.7	3.4	-2.7	19.6
Clerical	1.0	7.1	-6.0	19.9
Craftsman	3.5	6.5	-3.0	20.6
Sales	2.4	8.0	-5.6	22.7
Protective services	6.0	3.7	2.3	24.9
Arts, Entertainment, Writing	8.5	2.5	5.9	29.3
Teacher	8.2	5.5	2.7	32.5
Professional Medicine (not Doctor)	7.1	1.3	5.8	35.0
Other Professional	14.5	8.0	6.5	35.6
Engineer	11.0	6.2	4.8	38.7
Computer technical	3.8	9.2	-5.4	39.2
Manager	12.8	13.7	-0.9	41.3
Legal	4.8	1.2	3.6	53.0
Doctor	4.9	1.6	3.4	63.7
Military	0.7	1.5	-0.8	NA
Proprietor	6.5	5.8	0.7	NA
Observations	**2,306**	**2,410**		

Notes:
1 See notes to Table 8

Table 10. Proportion of 20 year old *workers* expecting to enter each profession by age 30, and the proportion who have actually entered that profession by age 26

	% expecting to be in the occupation at age 30	% actually in each occupation at age 26	Difference between expected and actual (% points)	Average CPS wage for each occupation in 2004 (deflated to 1994 $000)
Homemaker, not working, studying	0.6	3.9	-3.3	0.0
Farmer	2.3	0.9	1.3	12.5
Labourer	4.0	14.6	-10.6	16.7
Service	1.9	3.2	-1.3	17.0
Skilled operative	6.2	13.6	-7.4	19.6
Clerical	2.1	4.7	-2.6	19.9
Craftsman	16.6	19.4	-2.8	20.6
Sales	2.1	2.8	-0.6	22.7
Protective services	8.8	4.9	3.9	24.9
Arts, Entertainment, Writing	5.5	0.6	4.9	29.3
Teacher	3.2	0.6	2.5	32.5
Professional Medicine (not Doctor)	3.3	1.0	2.3	35.0
Other Professional	7.6	12.4	-4.7	35.6
Engineer	4.7	0.9	3.8	38.7
Computer technical	3.2	1.6	1.6	39.2
Manager	11.0	11.8	-0.8	41.3
Legal	1.6	1.2	0.4	53.0
Doctor	1.5	0.5	1.0	63.7
Military	1.4	1.2	0.2	NA
Proprietor	12.5	3.1	9.4	NA
Observations	**1,459**	**1,595**		

Notes:
1 See the notes to Table 8

Table 11. Ordinary least squares regression results comparing the accuracy of students' income expectations to workers

	Specification 1		Specification 2		Specification 3		Specification 4	
	Co	**SE**	**Co**	**SE**	**Co**	**SE**	**Co**	**SE**
Work-student status at age 20 (Ref: Working)								
Agriculture student	-0.27*	0.07	-0.20*	0.07	-0.21*	0.07	-0.29*	0.07
Economics, Finance student	-0.09	0.09	-0.02	0.09	-0.03	0.09	-0.14	0.09
Business, Management student	-0.03	0.05	0.02	0.05	0.00	0.06	-0.11*	0.06
Journalism, Communication student	0.18*	0.08	0.22*	0.08	0.21*	0.08	0.08	0.08
Computer Science, Maths student	-0.18*	0.09	-0.11	0.09	-0.15	0.10	-0.29*	0.09
Education student	-0.10	0.06	-0.05	0.06	-0.06	0.06	-0.15*	0.07
Engineering, Physical sciences student	-0.22*	0.05	-0.16*	0.06	-0.18*	0.06	-0.29*	0.06
Language student	-0.17	0.11	-0.10	0.11	-0.09	0.11	-0.22	0.12
Health student	0.13*	0.07	0.19*	0.07	0.17*	0.08	0.06	0.08
Law student	0.45*	0.19	0.45*	0.18	0.44*	0.18	0.26	0.17
Biological science student	0.20*	0.09	0.27*	0.09	0.27*	0.09	0.16	0.09
Social sciences, Humanities student	0.06	0.07	0.12	0.07	0.09	0.07	-0.03	0.07
Art student	0.26*	0.11	0.31*	0.11	0.31*	0.12	0.15	0.11
Other student	0.01	0.06	0.05	0.06	0.03	0.06	-0.10	0.06
Not student or working	0.12*	0.05	0.13*	0.05	0.11*	0.05	0.10*	0.05
Missing	0.28*	0.07	0.30*	0.07	0.26*	0.07	0.27*	0.07
Maths ability at age 18	-	-	-0.08*	0.02	-0.07*	0.02	-0.05*	0.02
Race (Ref: White)								
American Indian or Alaska Native	-	-	-	-	0.08	0.09	0.06	0.09
Asian or Pacific Islander	-	-	-	-	0.03	0.07	0.04	0.06
Black, not Hispanic	-	-	-	-	0.20*	0.05	0.18*	0.05
Hispanic or Latino	-	-	-	-	0.13*	0.05	0.11*	0.05
More than one race	-	-	-	-	0.09	0.09	0.07	0.09
Missing	-	-	-	-	0.07	0.08	0.03	0.07
Family income parents reported when respondent was age 18 (Ref: Bottom quintile)								
2nd quintile	-	-	-	-	-0.01	0.05	-0.01	0.05
3rd quintile	-	-	-	-	0.00	0.05	0.01	0.05
4th quintile	-	-	-	-	0.04	0.05	0.04	0.04
Top quintile	-	-	-	-	-0.01	0.06	0.02	0.06
Student at 20, who also held a part-time job (Ref: No)								
Yes	-	-	-	-	0.02	0.04	-0.03	0.04
College dropout (ref: No)								
Yes	-	-	-	-	-	-	0.31*	0.05
Constant	0.41*	0.03	0.38*	0.03	0.34*	0.05	0.36*	0.05

Notes:

1 The response variable is the natural logarithm of the ratio of expected to actual income

2 See notes to Table 1 for details on the "maths ability" variable

3 * indicates statistical significance at the 5% level

4 "Missing" dummy variables are included when the respondent has not provided information on any of the covariates. I do not present results for brevity.

Table 12. Robustness tests of accuracy of income expectations, using regression specification 3

	Test A		Test B		Test C		Test D	
	Co	S.E	Co	S.E	Co	S.E	Co	S.E
Work-student status at age 20 (Ref: Working)								
Agriculture student	-0.07	0.07	-0.11	0.07	-0.23*	0.07	-0.27	-
Economics, Finance student	-0.04	0.09	-0.08	0.11	-0.02	0.10	-0.03	-
Business, Management student	0.00	0.06	-0.01	0.06	-0.01	0.06	-0.02	-
Journalism, Communication student	0.24*	0.09	0.19*	0.08	0.20*	0.08	0.39	-
Computer Science, Maths student	-0.01	0.10	-0.13	0.1	-0.13	0.11	-0.10	-
Education student	0.10	0.07	-0.03	0.07	-0.06	0.07	-0.03	-
Engineering, Physical sciences student	-0.03	0.06	-0.17*	0.06	-0.19*	0.06	-0.14	-
Language student	0.02	0.12	-0.04	0.13	-0.08	0.11	-0.10	-
Health student	0.30*	0.07	0.19*	0.08	0.18*	0.07	0.17	-
Law student	0.51*	0.18	0.42*	0.20	0.48*	0.24	0.40	-
Biological science student	0.36*	0.08	0.25*	0.09	0.25*	0.09	0.43	-
Social sciences, Humanities student	0.23*	0.08	0.14*	0.07	0.11	0.07	0.17	-
Art student	0.39*	0.13	0.32*	0.14	0.29*	0.12	0.43	-
Other student	0.13*	0.06	0.03	0.06	0.02	0.06	0.13	-
Not student or working	0.16*	0.05	0.12*	0.05	0.13*	0.05	0.10	-
Missing	0.33*	0.07	0.27*	0.07	0.24*	0.07	0.42	-
Maths ability at age 18	-0.04*	0.02	-0.07*	0.02	-0.06*	0.02	-0.10	
Race (Ref: White)								
American Indian or Alaska Native	0.12	0.12	0.08	0.08	0.04	0.1	0.11	-
Asian or Pacific Islander	0.06	0.09	0.05	0.08	0.08	0.12	0.05	-
Black, not Hispanic	0.22*	0.06	0.22*	0.05	0.20*	0.05	0.16	-
Hispanic or Latino	0.13*	0.05	0.12*	0.05	0.14*	0.05	0.12	-
More than one race	0.09	0.07	0.16*	0.08	0.06	0.09	-0.06	-
Missing	0.16*	0.08	0.04	0.08	0.04	0.2	0.12	-
Family income parents reported when respondent was age 18 (Ref: Bottom quintile)								
2nd quintile	-0.09*	0.06	-0.03	0.05	-0.03	0.05	-0.01	-
3rd quintile	-0.02	0.05	-0.02	0.06	-0.02	0.05	-0.08	-
4th quintile	0.03	0.06	0.07	0.05	0.03	0.05	0.01	-
Top quintile	0.01	0.05	0.01	0.06	-0.03	0.06	0.03	-
Student at 20, who also held a part-time job (Ref: No)								
Yes	0.00	0.04	0.02	0.04	0.02	0.04	-0.04	-
Constant	0.51*	0.05	0.36*	0.05	0.37*	0.05	0.34	

Notes:
1 Test A refers to when I do not extrapolate the data, and simply compare age 26 income to expectations at age 30.
2 Test B refers to when I extrapolate the income data using prediction "Method 1" described in section 4.
3 In Test C, I re-weight the data to take into account the item non-response shown in Table 1 and 4. Robustness Test D presents the quantile (median) regression estimates. Note that standard errors have not been presented, due to the difficulties of providing accurate figures when using complex survey data (clustering and weighting) as in the NELS.
4 * Indicates statistical significance at the 5% level

Table 13. Logistic regression results comparing how realistic students' occupational expectations are to workers

	Specification 1		Specification 2		Specification 3		Specification 4	
	OR	**SE**	**OR**	**SE**	**OR**	**SE**	**OR**	**SE**
Work-student status at age 20 (Ref: Working)								
Agriculture student	0.61	0.34	0.52	0.28	0.59	0.35	0.74	0.45
Economics, Finance student	1.26	0.37	1.03	0.32	1.35	0.43	1.90*	0.60
Business, Management student	0.80	0.22	0.71	0.20	0.92	0.27	1.25	0.38
Journalism, Communication student	1.02	0.34	0.91	0.30	1.17	0.40	1.71	0.61
Computer Science, Maths student	1.53	0.63	1.27	0.50	1.72	0.67	2.82*	1.25
Education student	2.18*	0.57	1.92*	0.51	2.52*	0.73	3.41*	1.04
Engineering, Physical sciences student	1.86	0.44	1.59*	0.39	2.07*	0.55	2.83*	0.77
Language student	0.78	0.38	0.65	0.33	0.82	0.41	1.18	0.65
Health student	0.47┼	0.22	0.39*	0.19	0.56	0.27	0.78	0.38
Law student	1.01	0.41	1.01	0.40	1.40	0.55	2.56*	1.05
Biological science student	0.74	0.36	0.61	0.31	0.79	0.42	1.07	0.56
Social sciences, Humanities student	0.67	0.22	0.56┼	0.19	0.77	0.27	1.07	0.37
Art student	0.98	0.38	0.86	0.35	1.07	0.49	1.75	0.71
Other student	1.33	0.32	1.23	0.31	1.66*	0.43	2.51*	0.68
Not student or working	0.75	0.20	0.74	0.19	0.77	0.20	0.78	0.21
Missing	1.31	0.38	1.23	0.35	1.85	0.56	1.78*	0.55
Maths ability at age 18	-	-	1.20*	0.06	1.17*	0.06	1.11	0.07
Race (Ref: White)								
American Indian or Alaska Native	-	-	-	-	0.63	0.33	0.69	0.39
Asian or Pacific Islander	-	-	-	-	0.76	0.24	0.72	0.23
Black, not Hispanic	-	-	-	-	0.61	0.21	0.64	0.22
Hispanic or Latino	-	-	-	-	0.90	0.21	0.97	0.23
More than one race	-	-	-	-	0.85	0.29	0.89	0.32
Missing	-	-	-	-	0.60	0.24	0.70	0.27
Family income parents reported when respondent was age 18 (Ref: Bottom quintile)								
2nd quintile	-	-	-	-	0.90	0.21	0.93	0.21
3rd quintile	-	-	-	-	0.94	0.22	0.92	0.21
4th quintile	-	-	-	-	1.24	0.30	1.21	0.29
Top quintile	-	-	-	-	0.96	0.26	0.87	0.24
Student at 20, who also held a part-time job (Ref: No)								
Yes	-	-	-	-	0.58*	0.10	0.65*	0.11
College dropout (ref: No)								
Yes	-	-	-	-	-	-	0.30*	0.06

Notes:
1 * Indicates statistical significance at the 5 % level, ┼ Indicates statistical significance at the 10% level.
2 A higher odds ratio indicates more realistic occupational expectations.

Figure 1. Distribution of expected age 30 income (in 1994 US $) for young US males

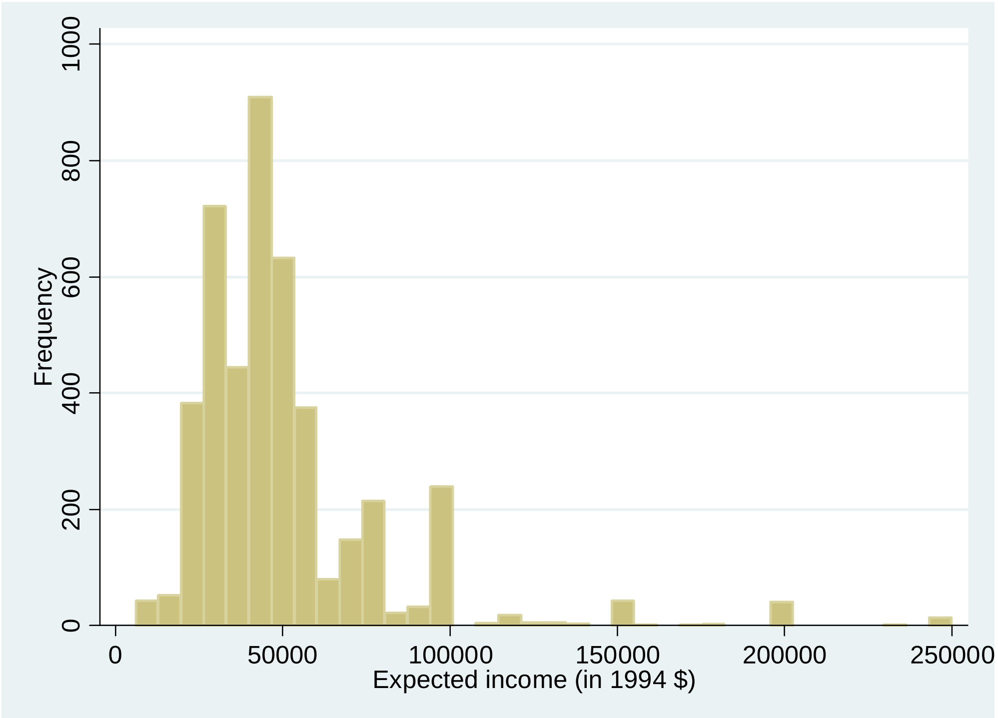


Figure 2a. Distribution of age 26 actual income (in 1994 US $) for young US males

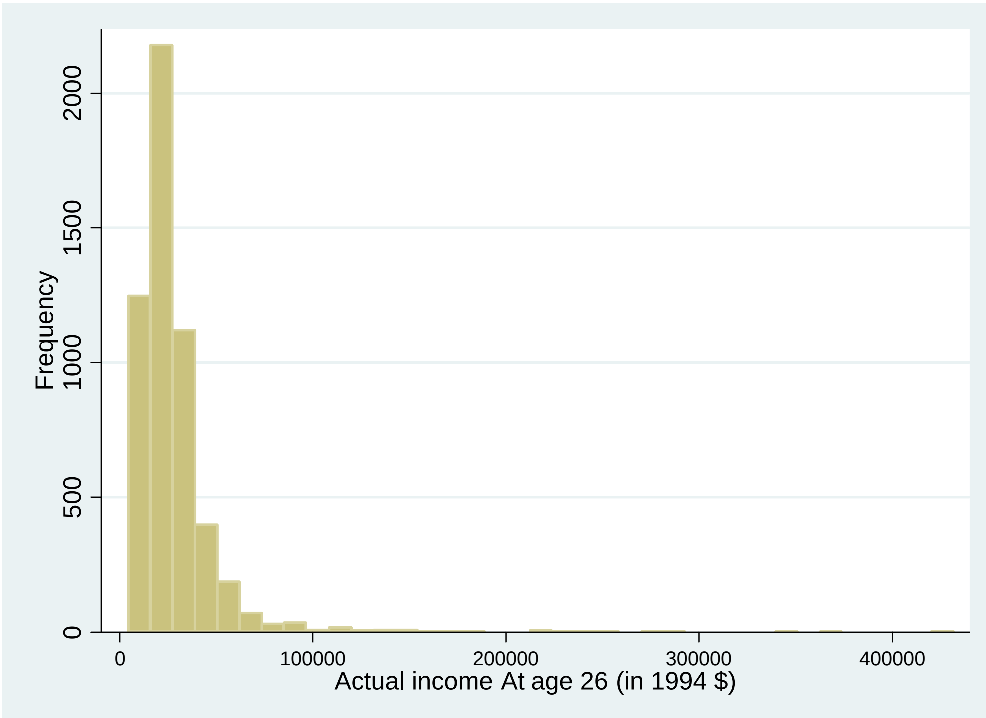


Figure 2b. Distribution of log age 26 actual income (in 1994 US $) for young US males

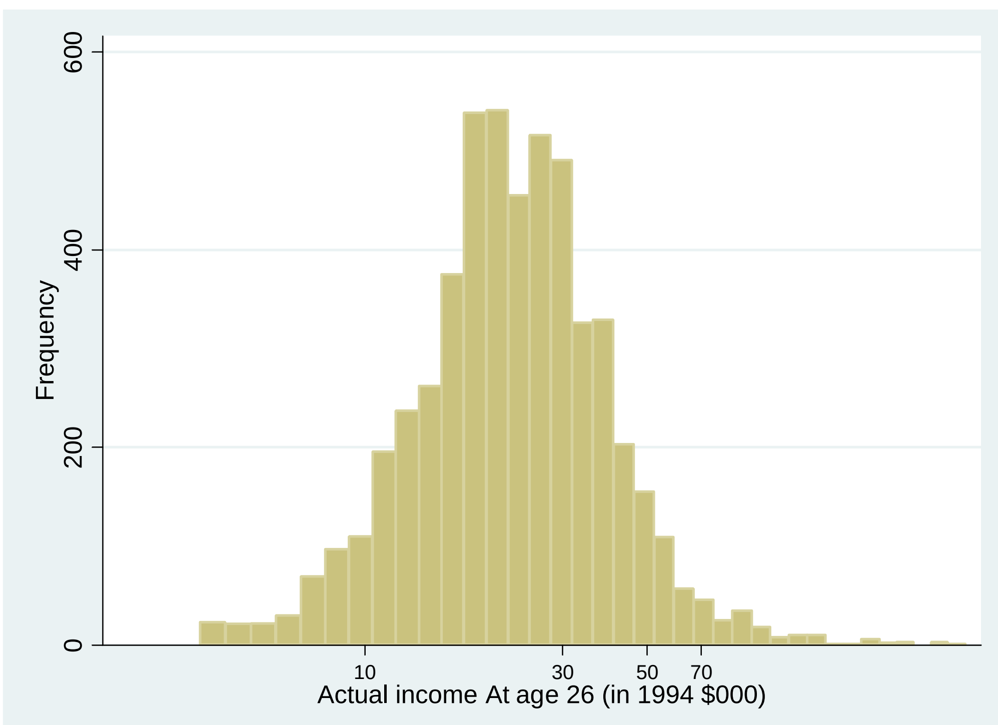
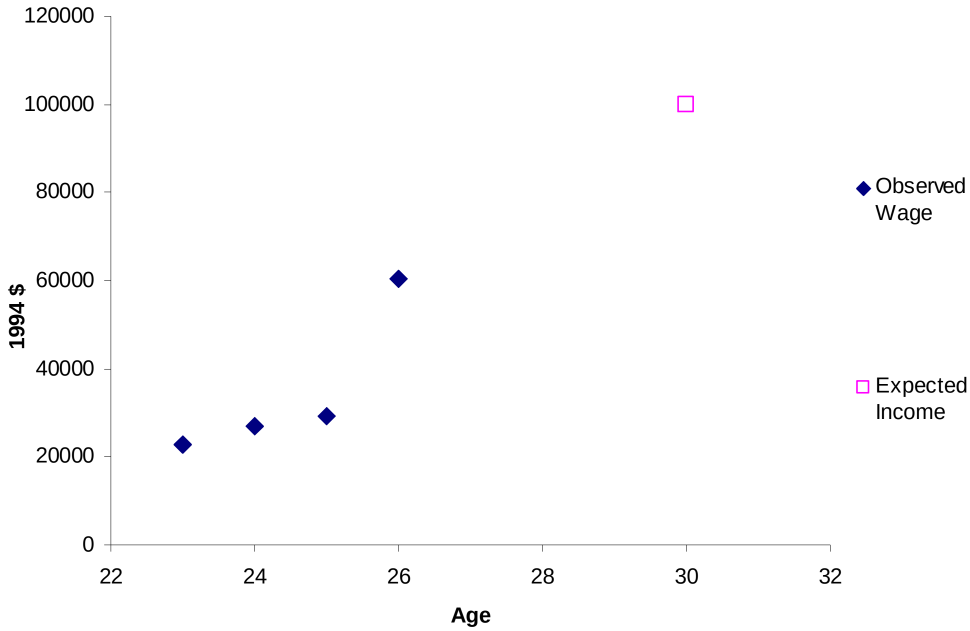
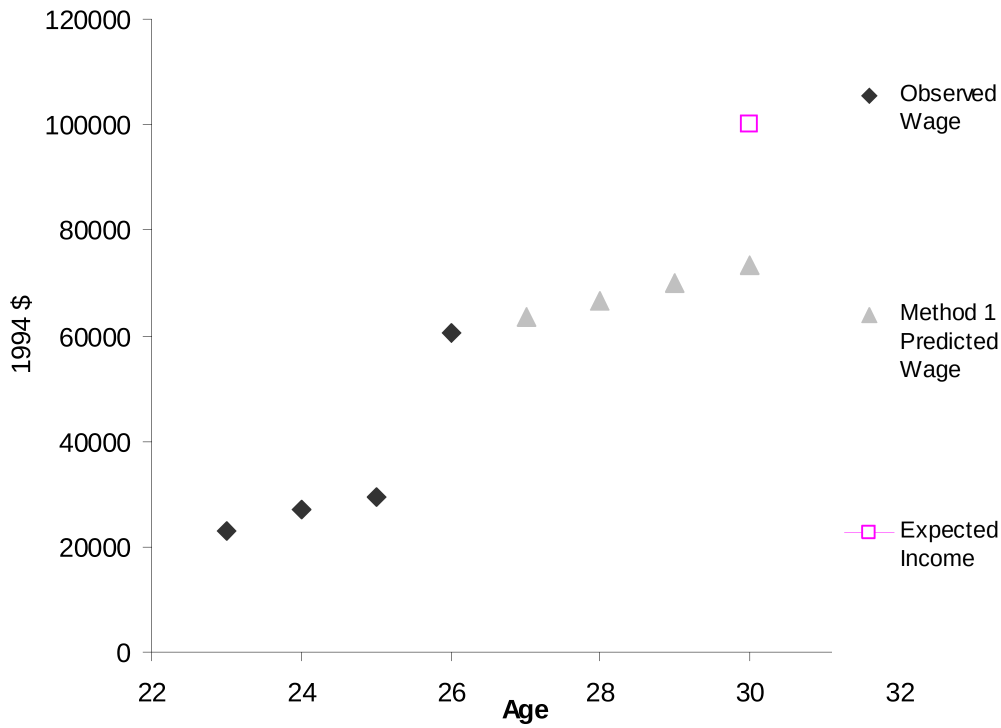


Figure 3. Data on expected income and actual wages that can be observed for one particular individual in the NELS



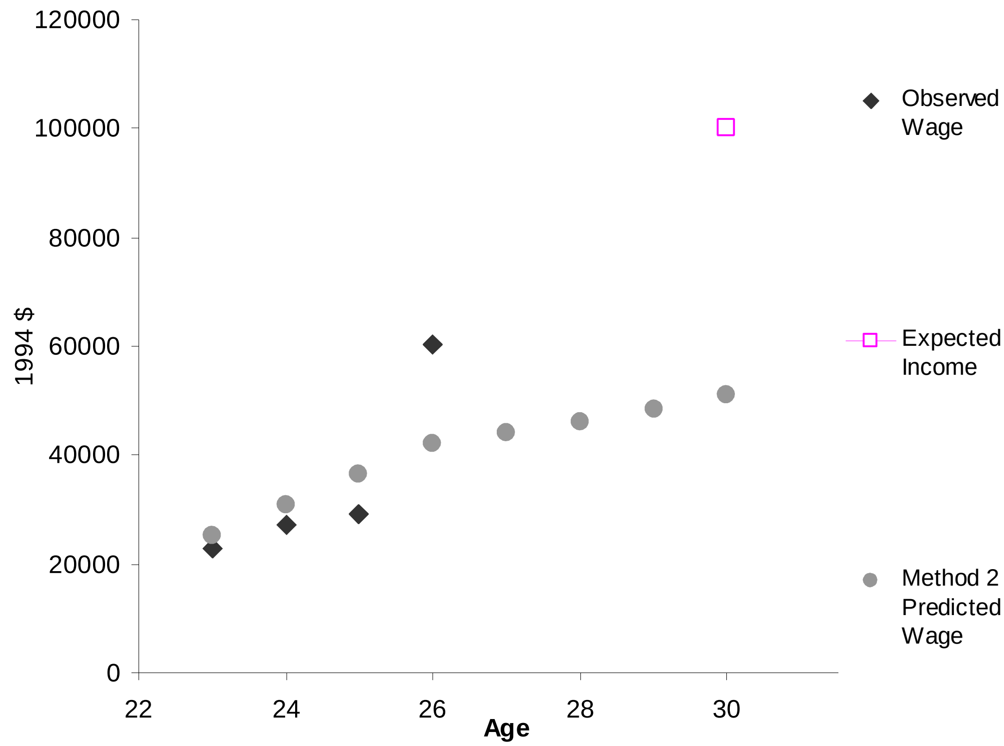
Note:
1 This individual reported zero unearned income at age 26; therefore his wages are equivalent to his income. Note that in these diagrams, I am simply trying to explain my extrapolation method for *wages*.
This individual is not an example of a "typical" NELS respondent. Rather, I have chosen this observation as it provides a good example of the points I am tying to make.

Figure 4. Illustration of wage prediction method 1 for ID 7286532 in the NELS



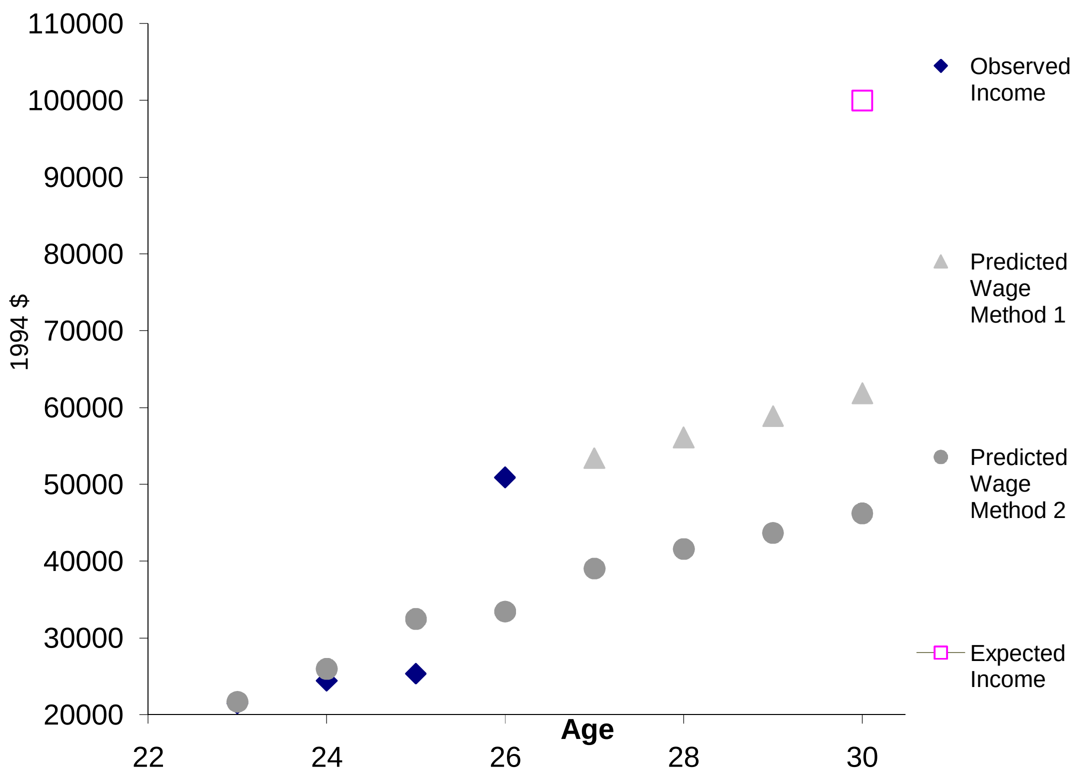
Note:
1 See notes to Figure 3
2 The above is a hypothetical example of extrapolation method 1.

Figure 5. Illustration of wage prediction method 2 for ID 7286532 in the NELS



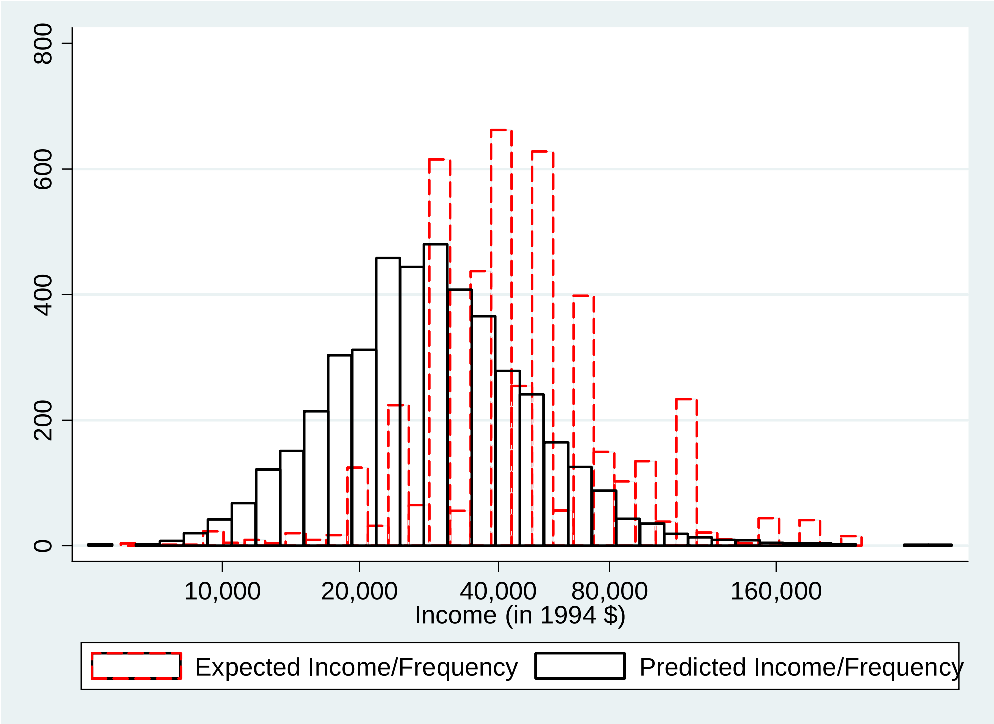
Note:
1 See notes to Figure 3
2 The above is a hypothetical example of extrapolation Method 2.

Figure 6. Comparison of wage prediction methods for ID 7286532 in the NELS



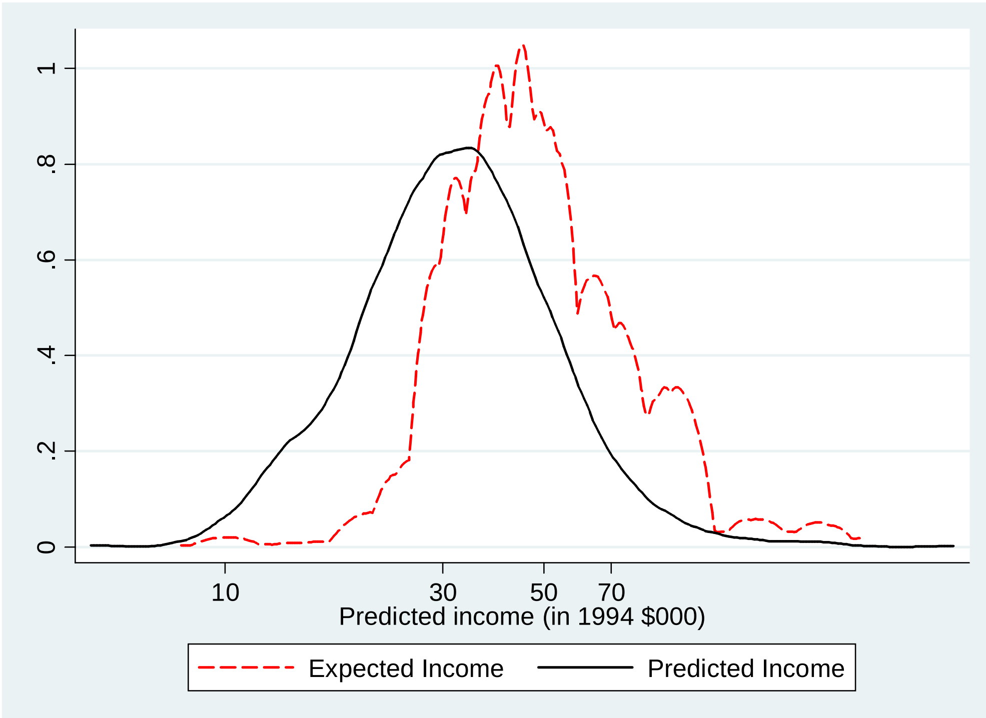
Note:
1 See notes to Figure 3, 4 and 5

Figure 7. Distribution of log expected and log predicted income at age 30



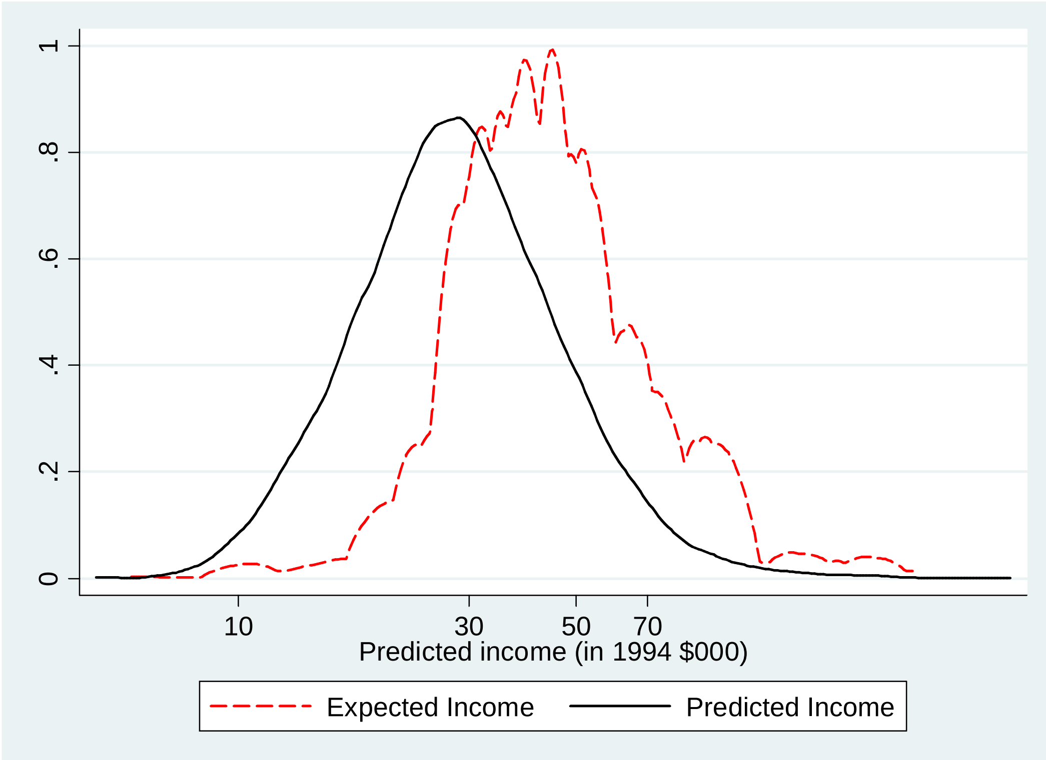
Note:
1 All data in 1994 prices
2 Predicted Income refers to that estimated using prediction method 2.
3 Dashed bars refers to distribution of expected income at age 30, solid bars refer to my predictions of actual income at age 30

Figure 8. Distribution of log expected and log predicted income at age 30 for students



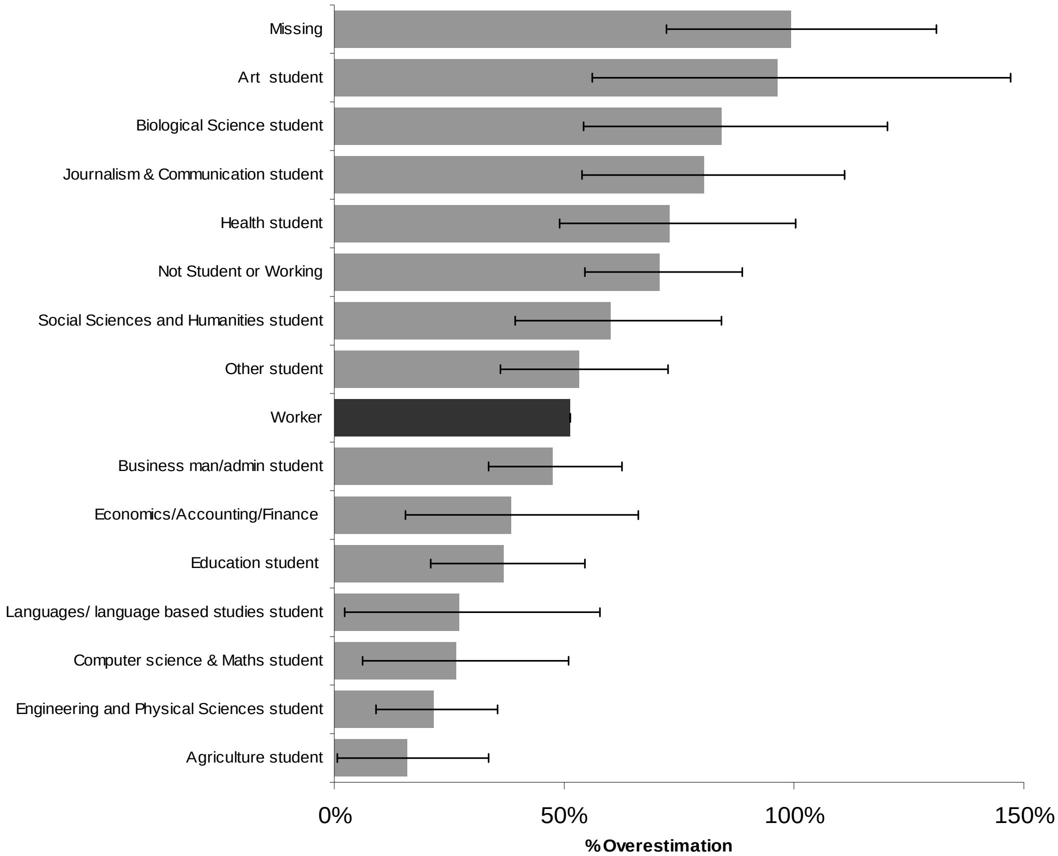
Note:
1 See notes to Figure 7

Figure 9. Distribution of log expected and log predicted income at age 30 for workers



Note:
1 See notes to Figure 7

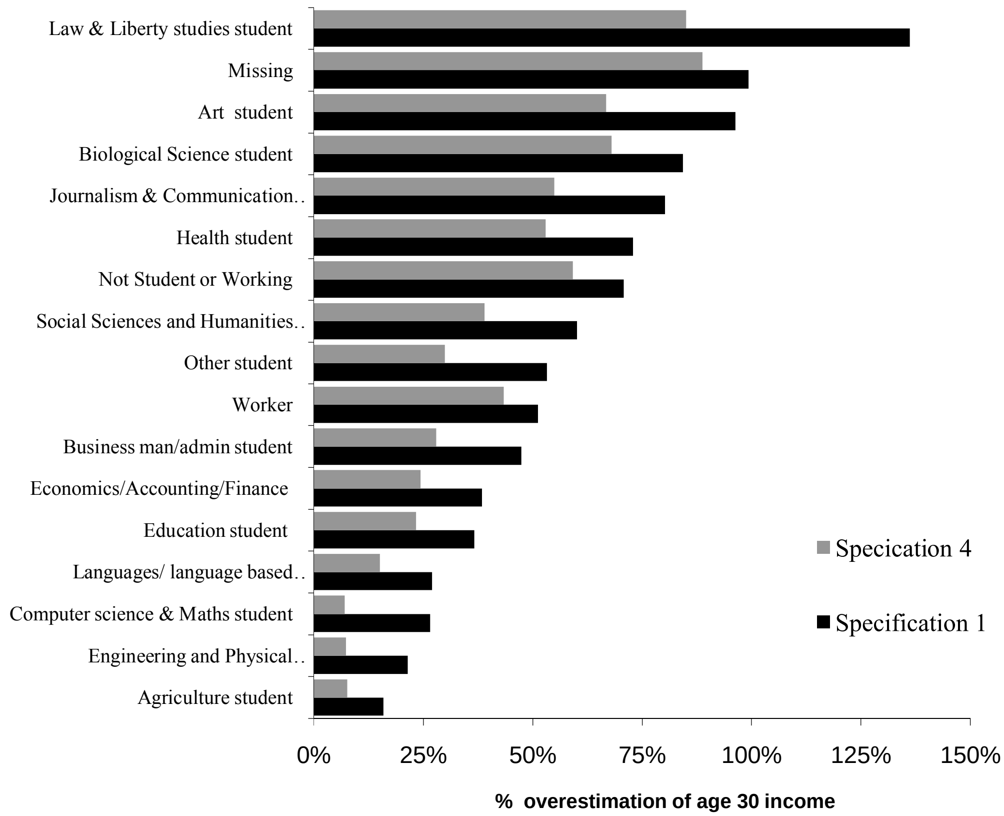
Figure 10. Difference between expectations and realisations: Workers compared to students in different subjects



Notes:

1 Thick bars refer to average overestimation of age 30 income for each group. These figures have been calculated from model specification 1, which just contains the subject dummy coefficients and no other explanatory variables. The thin black line running through the centre of each bar is the 95% confidence interval of this estimate.

Figure 11. Difference between expectations and realisations: Specifications 3 and 4



Notes:
1 Black bars refer to model specification 3, where I control for ethnicity, family income and ability in mathematics. The light grey bars refer to model specification 4, which also includes an indicator of whether the respondent had graduated from university by age 26.

APPENDIX 1.

NELS sample design[47] (Source: Curtin et al 2002)

This bulk of this Appendix has been taken (on occasion word for word) from Curtin et al (2002). These authors explain the NELS sample design at great length. I reproduce their work to help the reader understand some of the technicalities of the NELS sampling process. Although I have rephrased and edited part of their text, it should be noted that I claim none of Appendix 1 to be my own independent work.

The sample for NELS: 88/94 (i.e. age 20 sampling frame) was created by dividing the NELS:88/92 (i.e. age 18) sample into 18 groups based on their response history, dropout status, eligibility status, school sector type, race, test scores, socioeconomic status, and freshened status. Each sampling group was assigned an overall selection probability. Cases within a group were selected such that the overall group probability was met, but the probability of selection within the group was proportional to each sample member's second follow-up (age 18) design weight. Assigning selection probabilities proportional to the second follow-up (age 18) design weight, reduced the variability of the NELS:88/94 (age 20) raw weights and consequently increased the efficiency of the resulting sample from 40.1 percent to 44.0 percent. The groups were:

0. Excluded from age 20 follow-up

The age 20 follow-up sample is a spring defined sample. Therefore students who had been brought in through the freshening process, but who had dropped out by the time of data collection, as well as the age 14 dropouts were assigned to this group. As these groups have been excluded from the age 20 follow-up, they have a sampling probability of zero. In addition, sample members who were ineligible or out of scope (dead or out of country) for the age 18 follow-up were also assigned to this group.

1. Nonresponders

These sample members had never completed a questionnaire in any round

2. Poor responders

[47] See http://nces.ed.gov/pubs2002/2002323.pdf for more details

These are sample members who did not complete a age 18 questionnaire (but had responded at either age 14 or 16)

3. Ever dropped out

Sample members for whom Curtin et al (2002) have evidence that they ever dropped out of school (including those who were in school during periods of data collection) were included in this group.

4. Ineligible to participate (due to language barriers or mental or physical impairment) prior to age 18

5. Attended a private school at age 14

6. Attended a private school in either age 16 or 18

7. Hispanic

8. Asian or Pacific Islander (API)

9. Native American

10. Black, top quartile in cognitive tests

11. Black, other test scores

12. White, lowest socioeconomic quartile

13. White, highest socioeconomic quartile

14. White, middle socioeconomic quartiles

15. Freshened in at age 16

16. Freshened in at age 18

17. Other

The table below lists the groups, their selection probabilities and their age 16 and 18 follow-up distributions. While some sample members qualified for more than one of the sample groups, each member was assigned to only one group. The groups were created in order of priority, so that each sample member was assigned to the first group for which they qualified. For example, if someone was both a dropout (group 3) and was in a private school at age 14 (group 5), he or she was assigned to group 3.

The data used to assign the students to groups was drawn from a variety of possible sources, including questionnaire data for variables such as race and school sector type. If status at time of data collection was relevant and was not determined at the time of data collection, the imputed status developed during the age 18 weighting process was used.

Table A1. Sampling frame and selection probabilities NELS age 18 and 20 follow-up

	Selection Probability of being included in age 20 sample	N (Age 18 Sample)	N (Age 20 Sample)
TOTAL		21635	15964
Excluded	0	731	0
Non-responders	0.15	288	43
Poor responders	0.25	2383	596
Ever dropped out	1	2351	2351
Ineligible to participate	0.9	212	191
Attended private school at age 14	0.8	2984	2387
Attended private school at either age 16 or 18	0.8	122	98
Hispanic	0.9	1629	1466
Asian or Pacific Islander	1	874	874
Native American	1	132	132
Black, top quartile of cognitive tests	1	79	79
Black, other	0.9	1238	1114
White, lowest socio-economic group	1	1295	1295
White, highest socio-economic group	0.6	2536	1522
White, middle socio-economic group	0.8	4763	3810
Brought into sample at age 16	0.3	4	1
Brought into sample at age 18	0.3	6	2
Other	0.4	8	3

.

Table A2. Sampling frame and response rates – NELS age 20 follow-up

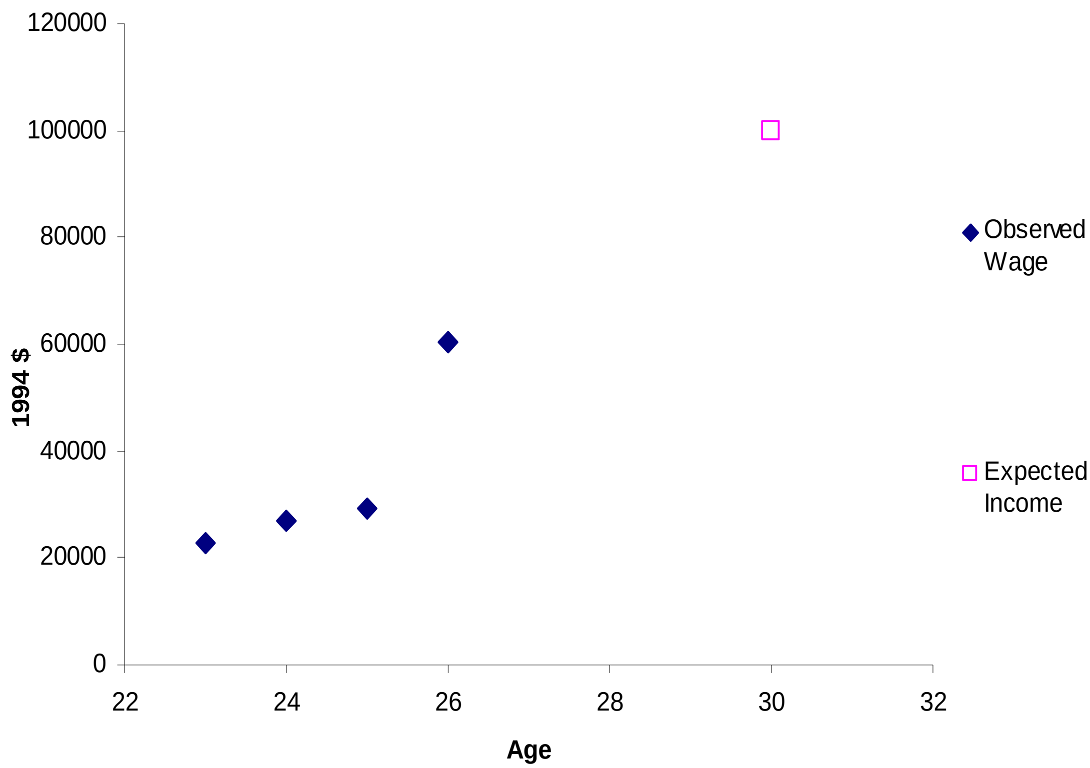
	Sampling frame age 20	Respondents to survey age 20
Gender		
Male	7,895	7,354
Female	7,980	7,561
Race/Ethnicity		
Asian	1,151	1,088
Hispanic	2,288	2,107
Black	1,840	1,681
White	10,303	9,787
Native American	230	211
Missing	63	41
Age 18 test quartile		
Lowest	2,669	2,497
2nd	2,850	2,710
3rd	2,836	2,746
4th	2,982	2,923
Missing	55	53
Did not complete test	4,483	3,986
Socio-economic status		
Lowest	4,062	3,788
2nd	3,784	3,587
3rd	3,742	3,570
4th	3,635	3,507
Missing	652	463
Drop out Status		
Never dropped Out	13,337	12,654
Ever dropped out	2,538	2,261
Age 14 school type		
Public	13,383	12,540
Catholic	1,355	1,292
NAIS private	595	568
Other private	542	515
Total	15,875	14,915

APPENDIX 2.

Methods to predict age 30 income

I now return to the two problems with the NELS data that I highlighted at the end of section 3, and briefly overviewed in section 4. To begin, consider Figure A1[48]. This illustrates the data observed for one particular individual in the NELS.

Figure A1. Observable wage and income expectation data for ID 7286532 in the NELS



Note:
All data in 1994 wages
This individual reported zero unearned income at age 26; therefore his wages are equivalent to his income. Note that in these diagrams, I am simply trying to explain my extrapolation method for *wages*. Discussion of unearned income can be found later in this Appendix.
This individual is not an example of a "typical" NELS respondent. Rather, I have chosen this observation as it provides a good example of the points I am tying to make. Most respondents see a gradual increase in their wage between 23 and 26, and not such a large increase at age 26.

[48] For this particular individual, the income they expect is significantly higher than their predicted income at age 30. This is not necessarily typical of all other respondents in the dataset. Rather I have chosen this individual as he is a good example of the substantial points I make throughout this section.

Respondents are asked what they expect their annual income to be when they turn 30. However, information on realisations is only available for wages between the ages of 23 and 26. Using the available data, I must make a prediction of each individual's age 30 income. I separate this into two parts: (a) the estimation of wages, and (b) the estimation of unearned income.

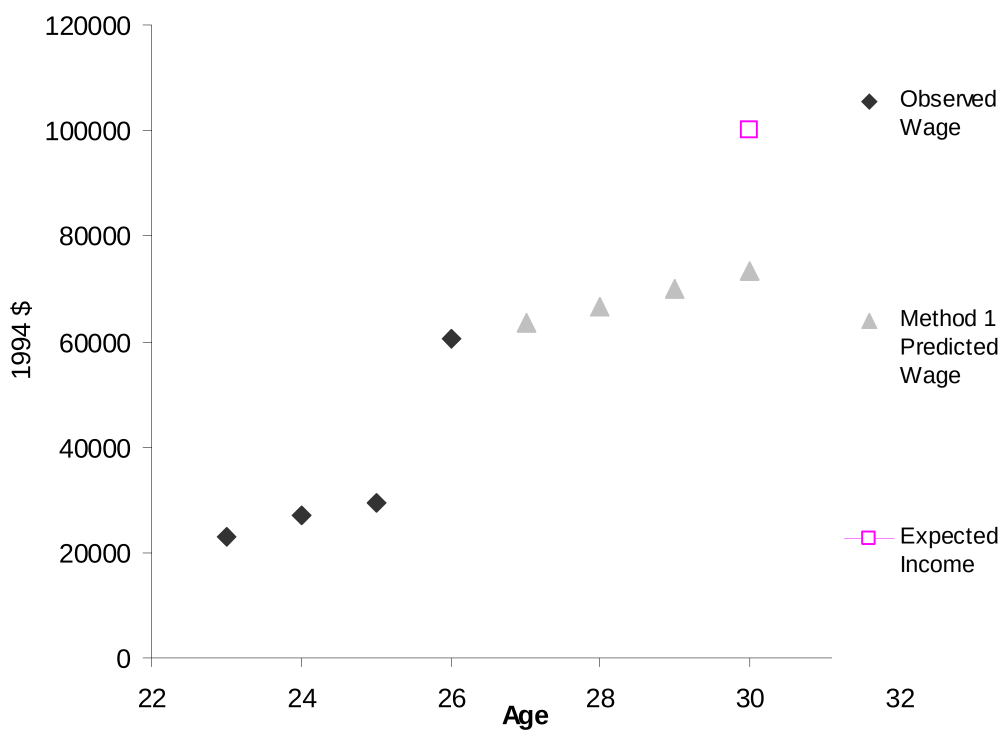
Wages

On average, wages grow quite substantially with the first ten years of labour market experience. Yet the path of wages for a given individual between 26 and 30 may be quite unstable. Both upwards and downwards shocks are possible due to job or career changes (promotion or redundancy), different family and location choices and preferences (prefer leisure to work due to the birth of a child) or simply macroeconomic conditions and "luck" (a particularly large bonus or commission in a given year). Given the factors described above, one may or may not wish to consider an individual's labour market history in predicting their age 30 wage. For instance, a sudden job change may make the information conveyed in past wages irrelevant. Recall Figure A1. At age 26, this individual has a particularly large wage by his "historical" (age 23- 25) standards. This may be the result of him changing job, perhaps into a much more lucrative career, where his labour market history is irrelevant for his future wages. Indeed, I do not know the reason for this potential change of job, but it could be a change in his preferences. For instance, this could be a graduate who took a low intensity job to enjoy life when young, though at age 26 made the decision to start a career. One may view this as a *permanent* shift in his wage profile. What he was previously earning, before this permanent shift, is irrelevant in predicting his future wage.

In contrast, it is equally possible for this to be a *temporary* increase in his wage, for instance from sheer good (or bad) luck. Take a 26 year old salesman who has had a particularly good year. For the salesman, the jump in wages could reflect a large bonus. However, in the future things may not be so good, with his wage reverting to his historical average (e.g. the average of the previous three years). Hence this observation at 26 may be treated more as an outlier, a sudden (but *temporary)* change in income.

Given the range of possibilities, I use two methods to predict age 30 wages. Method 1 views large changes in wages as a *permanent* shift in an individual's circumstance. Therefore, his previous wage profile is treated as irrelevant; it is only the most recently observed (e.g. age 26) wage that contains any useful information about his wage at age 30. Under this method, I simply take the most recently observed wage for each individual and extrapolate it forward, using external estimates of wage growth for young workers. Figure A2 presents a hypothetical example for the illustrative individual in Figure A1, assuming a real growth rate of 5% per annum. Observe that only the wage at age 26 influences my prediction, and that the large shock at age 26 gets carried forward. The previous income profile of the individual (between 23 and 25) has no influence at all.

Figure A2. Illustration of wage prediction method 1 for ID 7286532 in the NELS



Note: See notes to Figure A1

The above is a hypothetical example of extrapolation Method 1. I assume that his wage will grow at 5% per annum between the ages 26 and 30. His previous wage history (the wages he received between 23 and 25) play no part in the age 30 income prediction

To implement this method, I require an external estimate of the annual real wage growth for young workers. Rubinstein and Weiss (2007) provide a table of average annual real wage growth rates, as implied by a Mincer wage equation, broken down by labour market experience and educational attainment for three surveys; the Current Population Survey (CPS), Panel Survey of Income Dynamics (PSID) and National Longitudinal Survey of Youth 1979 (NLSY 79) [49]. Furthermore, Rubinstein and Weiss restrict each of the above datasets to full-time, male, American workers (as I have done with the NELS). One should note, however, that these surveys all relate to different years[50]. The growth rates they calculate from the CPS, PSID and NLSY are provided in Table A3, with further details available on page 14 and Appendix 5 of Rubinstein and Weiss (2007).

Table A3. Average, annual (real) wage growth rates for young workers: Rubenstein and Weiss estimates

		% Average (real) wage growth rate per annum by education level				
Number of years experience in the labour force	**Data source**	**Below high school**	**High school**	**Some college**	**College graduates**	**MA/PhD**
0-10	CPS	2.4	3.2	3.3	3.6	2.9
	PSID	2.8	3.0	3.8	3.9	3.2
	NLSY	2.4	3.4	4.6	5.2	5.5
11-15	CPS	1.6	2.2	-	-	-
	PSID	1.9	2.0	-	-	-
	NLSY	1.3	2.3	-	-	-

Notes:
Source: Table 1, page 14 of Rubinstein and Weiss (2007) Post Schooling Wage Growth: Investment, Search and Learning. Handbook of the Economics of Education, Volume 1

CPS: Current Population Survey Annual March Supplement
PSID: Panel Survey of Income Dynamics
NLSY: National Longitudinal Study of Youth 1979

[49] It should be noted that Rubinstein and Weiss provide two sets of growth rates, one based on a Mincer quadratic specification (experience and experience squared), and the other based on cell means. Their justification for the latter method is based on work by Murphy and Welch (1990), who claim the quadratic specification fits the age-earnings profile poorly, especially at the early stages of workers careers. One may worry that using the growth rates implied by a Mincer equation here could lead to underestimation of future wages. However, the paper by Murphy and Welch shows that the error in the quadratic wage specification is small after 3 years labour market experience and reaches zero at around 5 years. This means that for the period I am trying to extrapolate to, the quadratic Mincer specification fits the actual data quite well.

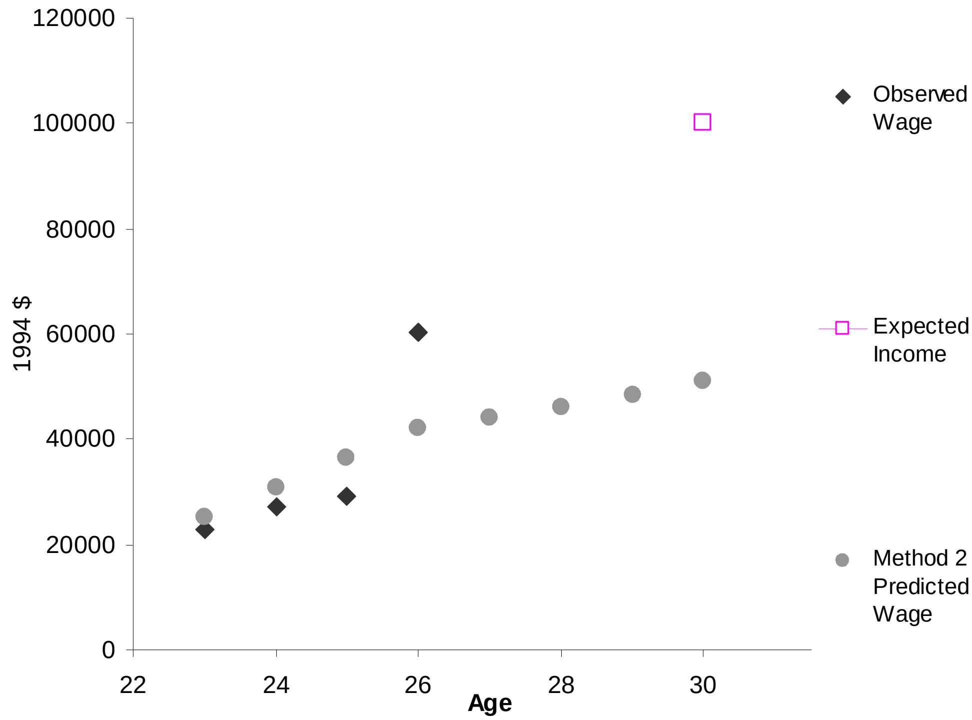
[50] The CPS data relates to wages between 1998 and 2002, the PSID is for all years after 1968, while the NLSY draws its information between 1979 and 2000.

These growth rates are applied to each individual in the NELS, depending on their highest qualification achieved by age 26. For example, an individual with college education, and who was earning $50,000 dollars at age 26, would be estimated to be earning $61,240 at age 30 (all in 1994 prices) [51]. In the event that wages go unobserved at age 26 (e.g. the individual was unemployed) I extrapolate from their last observed full-time wage[52]. From this point on, I shall call this prediction "Method 1"[53].

As suggested, this may not be an appropriate method if wage shocks (as for the respondent in Figure A1) are only temporary (e.g. a salesman with a large bonus). If this is the case, an individual's future wage will be randomly scattered around his time mean. The wage *history*, rather than just the most recent observation, is now informative. In prediction "Method 2" I take this into account. A hypothetical example is shown in Figure A3. To make the difference absolutely clear, contrast this with Figure A2 (that uses prediction Method 1). Even though both assume the same real growth rate (5% per annum) they generate quite different predictions of wages at age 30.

[51] $50,000 * ($1.052^4$), using the NLSY data and "College graduates" column in Table A3.
[52] For example, if someone was earning $50,000 at age 25, and their wage was not recorded at age 26, I would predict their age 30 income to be $50,000 * ($1.052^5$) = $64,577
[53] Rubenstein and Weiss provide growth estimates for young workers using three separate surveys (PSID, CPS and NLSY). I shall therefore use the notation "Method 1a" for my predictions when using their CPS growth estimates, "Method 1b" for their PSID estimates and "Method 1c" for their NLSY estimates.

Figure A3. Illustration of wage prediction method 2 for ID 7286532 in the NELS



Note:
See notes to Figure A1
The above is a hypothetical example of extrapolation Method 2.

Hence "Method 2" uses individual's wage history, rather than just the most recent observation, to predict future income. To implement Method 2 I use a fixed effects regression model, following the methodology of Carneiro and Heckman (2003)[54]. The natural logarithm of wages is the dependent variable, with age and age-squared as (time varying) explanatory terms that capture wage growth[55]. Other specifications, such as including a cubic age term or using a set of age dummies as an alternative, were also estimated. However results did not differ significantly from the parsimonious quadratic specification. I run separate regressions for five different educational groups, based on highest qualification achieved by age 26[56].

[54] Carneiro and Heckman (2003) faced a similar problem in having to estimate wages up to age 65 for a sample whose last observed wage was at age 35. In particular, they pool their data with an additional source, and use a similarly specified fixed effects model to estimate wages into the future. However, unlike Carneiro and Heckman, I run separate regressions for each educational group. Moreover, they specify an autoregressive error term, whereas I assume it to be random. In other words, they allow the particularly large wage shown in Figure A3 to revert to the mean (estimated fixed effect) over a series of years, whereas I assume it returns there instantly in the next period. Hence the method of Carneiro and Heckman is a sort of middle ground between the two extrapolation methods that I am proposing here. I also experimented with an auto-regressive error term, but found no change to my substantial results.

[55] Age enters as a quadratic term to allow for flattening of the age-earnings wage profile.

[56] Minicozzi (2003) suggests using separate regressions for different education-occupation combinations. Here, separate regressions are estimated only for different educational groups. Considering occupation would have

This allows the age coefficients, and therefore wage growth, to vary between groups with different levels of human capital. I have also experimented with alternative specifications that allowed wage growth to vary within these educational groups, and found similar results[57]. In all models, I assume the error term is independent and identically distributed, scattered randomly around each individual's fixed effect. Formally, this model can be expressed as:

$$Y_{ia} = \alpha + \beta_0 A_{ia} + \beta_1 A_{ia}^2 + \eta_i + \varepsilon_{ia} \; \nabla \in E$$

Where:

Y_{ia} = log earnings of individual i at time a

A_{ia} = Age of individual i at time a

η_i = Individual (or fixed) effect

ε_{ia} = Error term, assumed to be normally distributed

E = Five education groups (Less than high school, high school, associates degree, bachelors degree, MSc/PhD)

In this model, it is the estimated fixed effect, η_i, that captures the influence of all wages for individual i between 23 and 26. Note that this specification is quite different to a "standard" wage equation, where the aim is to estimate the impact of various regression coefficients on the outcome (wages). My concern, on the other hand, is not in estimating the importance or effect sizes of various explanatory variables, but in predicting future wages. Therefore I allow the individual fixed effect to capture all the factors that are usually included on the right hand side of "standard" wage equation. This includes geographic location, individual ability and socio-economic background[58].

The estimated coefficients from the five regressions enter a prediction equation for age 30 wages, formally specified as:

lead to vastly reduced sample sizes and imprecisely estimated coefficients.

[57] In particular, I estimate a model where I allow wage growth to differ between college students who study different subjects. All the results presented in section 5 are robust to these additional specifications.

[58] However, I do account for human capital separately by estimating five regressions based on each individual's highest educational attainment at age 26.

$$\hat{Y}_{i30} = \alpha + \hat{\beta}_0 A_{i30} + \hat{\beta}_1 A_{i30}^2 + \hat{\eta}_i + \hat{\varepsilon}_{i30} \qquad \nabla \in E$$

With

$\hat{Y}$ = Predicted log wage at age 30

A= Age

$\hat{\eta}$ = Individual fixed effect

$\hat{\varepsilon}_{30}$= Random draw from the distribution of errors at age 26 (assumed to be normally distributed)

E = Achieved education at age 26

This prediction includes an error term. I assume the errors at age 30 are normally distributed with mean zero and variance equal to that in the estimated error distribution at age 26. I then take a random draw from this normal distribution for each individual.

One could estimate the preceding model, and form predictions, based solely on the NELS data. However this would have some fairly significant disadvantages. The age coefficients, which reflect wage growth, would be estimated solely from data in the observed period (wages recorded between the ages of 23 and 26). One would be assuming that the annual wage growth rate between 26 and 30 is the same as the wage growth rate between 23 and 26. This seems unlikely. Murphy and Welch (1990) show that earnings between 23 and 26 grow substantially faster than between 26 and 30. Moreover, with wages recorded at only 4 time points in the NELS, the quadratic age function would be poorly defined. On the other hand, using just a linear age function would miss an important empirical feature (flattening) of the age-earnings profile.

Thus the NELS data must be complemented with additional information on how wages grow in the unobserved period (27 to 30 years old). One method, used by Carneiro and Heckman (2003), is to pool the truncated survey (the NELS, which only contains wages until 26), with a second comparable data source that follows individuals to the point of interest (up to age 30). This pooled dataset will therefore contain information on wages between 23 and 30. However, certain criteria must be checked and some assumptions must be made. In particular, one implicit assumption is that the (unobserved) wage growth rate experienced by NELS sample members between the ages of 26 and 30 will be the same as the (observed) growth rate experienced by sample members from the second pooling survey. A further assumption is that structural differences in the economy, and between the two samples, do not lead to differences in wage *growth*[59]. It is also vital the two surveys are collecting comparable data, with similar wording of key questions.

The survey chosen is the National Longitudinal Study of Youth (NLSY) 1979[60]. The survey began in 1979, with 12,686 men and women surveyed who were between the ages of 14 and 22. These individuals were then followed each year, and have information on their income at age 30 collected between 1987 and 1995. I make similar restrictions in the NLSY as I have done in the NELS (I have excluded women and only consider the wages of individuals when they are working full-time).

The NLSY has numerous attractions as a source to pool with the NELS. Critically, wages are collected for individuals between the ages of 23 and 30, providing information on wage growth during the period not observed (between 26 and 30) in the NELS. Secondly, the wording of the questions regarding wages is broadly similar. Whereas the NELS asks:

First, including all of the wages, salaries, and commissions you earned in (1997/1998/1999), about how much did you earn from employment before taxes and all other deductions?

The NLSY uses the similar phrase:

Not counting any money you received from your military service during (YEAR), how much did you receive from wages, salaries, commissions or tips from all jobs, before deductions for taxes or anything else?

---

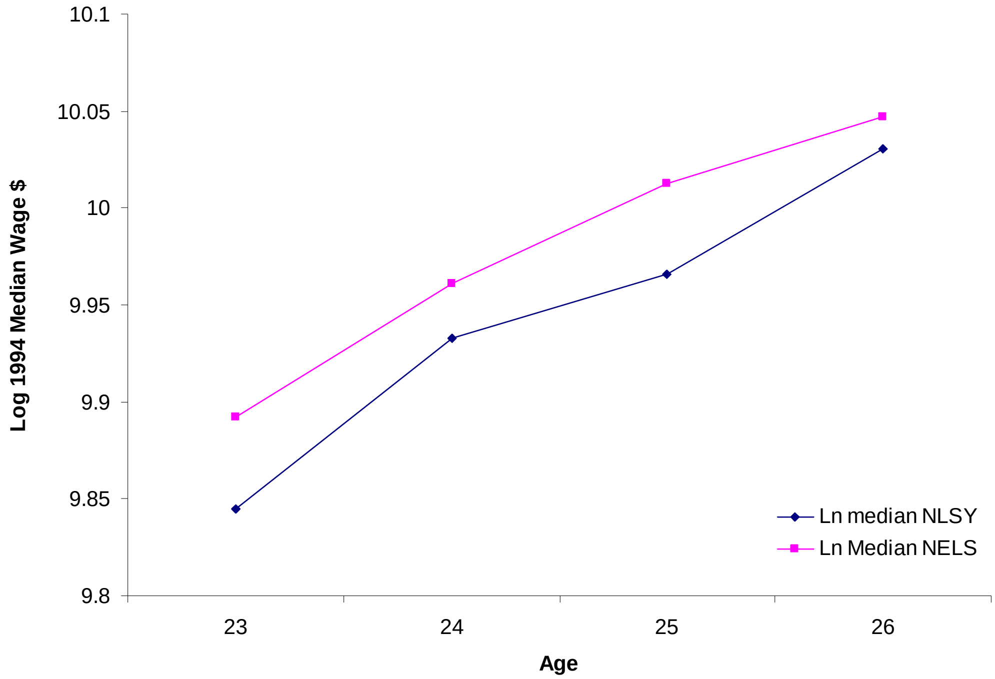
[59] Of course, overall wage levels are likely to be higher for cohorts from later periods. However this general rise in the wage level should be captured by the person specific fixed effect.

[60] The United States Department of Labor describes "The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14-22 years old when they were first surveyed in 1979". The National Bureau of Economic Research describe the quality of this survey, and the high response rate. In particular, it notes that 87% of those selected for interview responded in the base year (1979), while 86.7% of eligible respondents took part in 1996. Further details can be found at www.nber.org/~kling/surveys/NLSY79.html and http://www.nlsinfo.org/nlsy79/docs/79html/79text/front.htm

Although the NLSY excludes military wages in this question, this has been recorded separately in another item. Thus this is easily added in to also include those who were working in the military. Otherwise, the comparability seems good. Both ask for gross wages, before tax or other deductions. Moreover, it is made clear that the respondent should be taking into account all aspects of employment related income, including all wages and commissions, from all employment held.

It is also important to check that during the observable period (ages 23 to 26), wages in the two surveys grow at a similar rate. If wage growth is vastly different during the period observed in both surveys, it would be difficult to justify the assumption that the NELS will have similar wage growth to the NLSY in the unobserved (ages 27 to 30) period. Appendix Figure A4 presents the median wage recorded at each age in the NELS and NLSY. Although the median wage in the NELS is above that in the NLSY, the growth in wages between 23 and 26 is similar. For instance, mean (median) wages grew by 23% (17%) in the NELS between the ages 23 and 26, compared to 23% (22%) in the NLSY. In terms of the mean, this difference is very small. Though the gap is larger for the medians (around 1% per year), it still seems that the growth rate in the two surveys is reasonably similar. This, hopefully, means the wages that go unobserved in the NELS, between 27 and 30, will also follow a similar growth pattern to the NLSY.

Figure A4. Log median wages in the NELS and NLSY between the ages 23 and 26



To further investigate this point, I conduct a Chow test to investigate whether the age coefficient (which reflects wage growth) differs between the NELS and NLSY for wages observed between 23 and 26. Since wages are only available for four time points, age is kept as a linear function. Separate regressions are ran for each education group, and takes the form:

$$Y_{ia} = \alpha + \beta_0 A_{ia} + \beta_1 A_{ia} * S_i + \eta_i + \varepsilon_{ia} \quad \nabla \in E$$

Where:

Y_{ia} = log earnings of individual i at time a

A_{ia} = Age of individual i at time a

S_i = Dummy variable indicating respondent was part of the NELS survey (NLSY reference group)

η_i = Individual (fixed) effect

ε_{ia} = Random error term

E = One of five education groups (Less than high school, high school, associates degree, bachelors degree, MA/PhD)

The test is formally specified as:

H_O: $\boldsymbol{\beta}_1 = 0$

H_A: $\boldsymbol{\beta}_1 \neq 0$

The results are given in the Table A4. As expected, all age terms are significant. However the real interest rests on the Age-Survey interaction terms (the column labelled "Difference"). These show whether growth in average wages, for each schooling group, differs between the NELS and NLSY between the ages of 23 and 26. Out of the five regressions, two are statistically significant. However, these are for the two smallest education groups, which in total make up only around 10% of the NELS sample. Indeed the coefficient for MSc/PhD graduates is likely to be poorly defined due to the limited number of wage observations in the NELS for these individuals. In the other three regressions, which account for 90% of observations, the age-survey coefficient is very small. Indeed, if the Chow test is performed on the sample as a whole, with no distinction between education groups, the results suggest a difference in wage growth rates of under 1% per year. This is consistent with my claim that the NELS and NLSY cohorts experience similar wage growth patterns between 23 and 26 years of age. The assumption, based on this result, is that the NELS and NLSY samples also experience similar growth rates between 26 and 30.

Table A4. Chow test to investigate whether wage growth is similar for young men between the ages 23 and 26 for the NELS and NLSY surveys

Education group	Average per annum real wage growth NLSY	Average per annum real wage growth NELS	Difference
Below high school	0.049 (0.0055)	0.047	-0.002 (0.0128)
High school graduate	0.057 (0.003)	0.051	-0.006 (0.006)
Associates degree	0.093 (0.0061)	0.072	-0.021 (0.010)*
Bachelors degree	0.145 (0.0067)	0.141	-0.004 (0.008)
MSc/PhD	0.155 (0.0155)	0.250	0.095 (0.025)*

Notes:
The difference column relates to the Chow test of whether wage growth differs between the ages of 23 and 26 for each education group in the two surveys. In other words, this is the test of Ho: $\beta_0 = 0$ in the hypothesis test specified above.
Standard errors are presented in parenthesis
* Indicates statistical significant at the 5% level

A final issue may be that structural changes in the economy led to differential growth rates in real wages between the two surveys. However between 2000 and 2004 (the period unobserved in the NELS) average annual real wage growth in the US was around 3.2%. Between 1987 and 1995 (when members of the NLSY were turning 30) wage growth averaged around 4%. Although there is a difference, it appears to not be substantial, though it could lead to a slight overestimation of the predicted age 30 NELS wage.

Having established the comparability of the two surveys, I proceed to pool the information from these two datasets together (which, from this point on, I will call the NELS-NLSY pooled data). Using this data, I predict age 30 wages by estimating the fixed effect regression specified on pages 11-12 of this web Appendix. Table A5 provides the regression coefficients for the age and age squared terms from the five estimated models (recall I estimate five separate regressions based on educational attainment by age 26).

Table A5. Estimated age coefficients from prediction method 2 (fixed effects regression model)

Education level at age 26	Variable	Coefficient	SE
Below high school	Age	0.051	0.0071
	Age^2	-0.003	0.0010
High school	Age	0.053	0.0037
	Age^2	-0.003	0.0005
Associates degree	Age	0.091	0.0072
	Age^2	-0.004	0.0010
Bachelors	Age	0.171	0.0051
	Age^2	-0.011	0.0008
MA/PhD	Age	0.200	0.0152
	Age^2	-0.011	0.0020

Notes:
These are the estimated age coefficients from the fixed effects regression model described above. These coefficients reflect the estimated wage growth between 26 and 30. Table A6 converts these coefficients into estimated annual wage growth for ease of interpretation and comparison to the wage growth rates suggested by Rubenstein and Weiss.
All coefficients are statistically significant at the 5% level

All coefficients are statistically significant at the 5% level, capturing the quadratic effect of age and the flattening of the age-earnings profile. For interpretation purposes, however, it is easier to convert these coefficients into estimated annual growth rates (as per Rubinstein and Weiss). I present these in the final column of Table A6, which also contains the average annual wage growth estimates from prediction Method 1 for comparison.

Table A6. Predicted average, annual real wage growth rates for young American men between the ages 26 and 30

	Estimated % real growth rate per annum			
	Method 1			Method 2
	(a) CPS	(b) PSID	(c) NLSY	NELS-NLSY Pooled
Below high school	1.6	1.9	1.3	1.6
High school	2.7	3.6	2.8	2.5
Associates degree	3.3	3.8	4.6	5.2
Bachelors	3.6	3.9	5.2	5.0
MA/PhD	2.9	3.2	5.5	6.6

Notes:
"Below high school" annual growth rate is taken from the 11-15 years experience row in Rubenstein and Weiss, under the assumption that they would have (potentially) entered the labour market at 16/17.
For annual growth rates of high school graduates, I use a simple average of the Rubenstein and Weiss 0-10 years experience and 11-15 years experience columns. The assumption is that high school graduates enter the labour market at 18, and thus have around 8 years labour market experience at age 26 increasing to 12 years experience at age 30.
Individuals in all other educational groups (associates degree, bachelors degree and MA/PhD) are assumed to have 0-10 years labour market experience between 26 and 30.

It appears predicted wage growth is similar across all methods for each of the educational groups. As expected, the estimated growth rates from Method 2 are the best aligned with Method 1c (Rubenstein and Weiss's estimated wage growth when they use the NLSY data), though are slightly higher (lower) for the top (bottom) educational group[61].

I also present the average predicted age 30 wage for the two methods in Table A7. Predicted age 30 wages are similar across estimation methods, with differences typically less than 5% for both the mean and the median. However, the spread of Method 2 (the fixed effect model) is smaller. Indeed this is as expected; outlying observations get moderated in Method 2 by the influence of previous wages (it is a time mean). This does not occur in Method 1, where it is only the most recent observation that is used for prediction. Hence if there is a large shock to the most recent observation, this gets carried forward to the future prediction in Method 1, as opposed to being averaged out in Method 2.

[61] The Rubenstein and Weiss growth rate "Method 1" is based upon an average for the first ten years labour market experience. On the other hand, in the NELS, those with MSc/PhD level education are only likely to have up to 5 years experience between 26 and 30. Likewise, those with high school education in the NELS will have between 8-12 years experience. This is the most likely reason for the slightly differences, and that Method 1 growth rates are slight under (over) estimates for the most (least) educated.

Table A7. Predicted annual wage at age 30 for NELS sample members

Prediction method	Dataset	Median predicted wage $000	Mean predicted wage $000 (standard deviation)
1	CPS	25.7 (3.4)	29.4 (19.2)
1	PSID	25.5 (3.4)	29.5 (19.3)
1	NLSY	26.1 (3.4)	30.4 (20.1)
2	NELS-NLSY Pooled	26.7 (3.3)	29.6 (14.8)

Notes:
All figures presented in 1994 wages
Figures in parenthesis represent the spread of the data (p90/p10 for median, standard deviation for the mean)

I also compare my predictions of average age 30 wages for different groups to similar information recorded for 30 year olds in an external data source (the 2003-2005 CPS March Annual Supplement)[62]. The results appear in Table A8.

[62] The exact wording in the CPS is as follows: "How much did (name/you) earn from this employer before taxes and other deductions during (Year)?". This is supplemented with other questions to check the robustness of answers and to calculate other wage sources. In particular, respondents are asked "How much did (name/you) earn in tips, bonuses, overtime pay or commissions from this employer in (Year)" and "What is your best estimate of (name's/your) correct total amount of earnings from all other employers during (Year)?". All of these responses are used to calculate respondents earned income, making the definition comparable to the other surveys in question (NELS, NLSY, PSID).

Table A8. Predicted *mean* age 30 NELS wage compared to the *mean* age 30 Current Population Survey (CPS) wage

	% of observations in NELS	Predicted wage method 1c ($000)	Predicted wage method 2 ($000)	% of observations in CPS	CPS wage ($000)
Highest qualification at age 26					
Below high School	5.4	20.9	20.8	12.4	15.8
High school	56.4	27.0	25.5	47.8	24.7
Associates degree	7.0	29.9	30.5	8.2	28.1
Bachelors	28.0	37.7	37.8	24.6	38.1
Masters degree / PhD	2.8	42.4	43.6	6.8	44.5
Race					
White	69.8	31.4	31.3	60.6	31.5
American Indian	0.1	23.6	24.6	0.2	NA
Asian or Pacific Islander	5.1	38.5	36.4	6.2	35.4
Black (not Hispanic)	8.3	24.6	25.1	9.6	26.9
Hispanic	13.3	27.4	27.1	21.7	20.7
Other	4.6	27.1	26.6	1.7	NA
All respondents	100.0	30.4	29.6	100.0	28.9

Notes:
All observations in 1994 $
Total sample size in the NELS is 4,434. In CPS, the total sample size is 1,412
CPS Wages for American Indian and Other ethnic groups are not reported due to the small sample size.
CPS data from 2002-2004 March Annual Supplement, restricted to men working full-time, all year round, at age 30 available from : http://www.census.gov/hhes/www/cpstc/cps_table_creator.html
The proportion of respondents with below high school education is lower in the NELS than CPS. This is because CPS is a general population survey. The NELS data I am using represents the population who were in high school as seniors in 1992. Hence the NELS and CPS cover slightly different populations, particularly regarding those with less than high school education. On the other hand, the CPS contains a lot more individuals with MA or PhD level qualifications. This is due to the last NELS wave being conducted at age 26. Some of the NELS cohort would still be in higher education, and are still studying for these qualifications.

In general, predicted age 30 wages are similar to those in the external CPS data. I predict average wages to be $29,600 in the NELS, while in the CPS the equivalent figure is $28,900. Even when looking at subgroups of the population, differences tend to be quite small. For example, I predict the mean wage of those with a Bachelors degree to be $37,800, while in the CPS the average wage for those with a degree is $38,100. Likewise, I predict White respondents to earn a mean wage of $31,300, while in the CPS the figure is $31,500. Nevertheless, there are some groups where predicted wages are quite different from average wages in the CPS. For instance, I estimate the average wage of those with less than high school education to be around $21,000 while the CPS figure stands at just over $15,000. This may be because the NELS data represents the population of high school seniors in the spring of 1992. Hence my definition of "less than high school education" is those who made it into the final year of high school but did not graduate. On the other hand the CPS represents the whole US population, and defines less than high school education as everyone who did not graduate from high school, *including* those who dropped out *before* their senior year. This is probably the reason why, in the NELS compared to the CPS:

(a) my predicted wage is higher

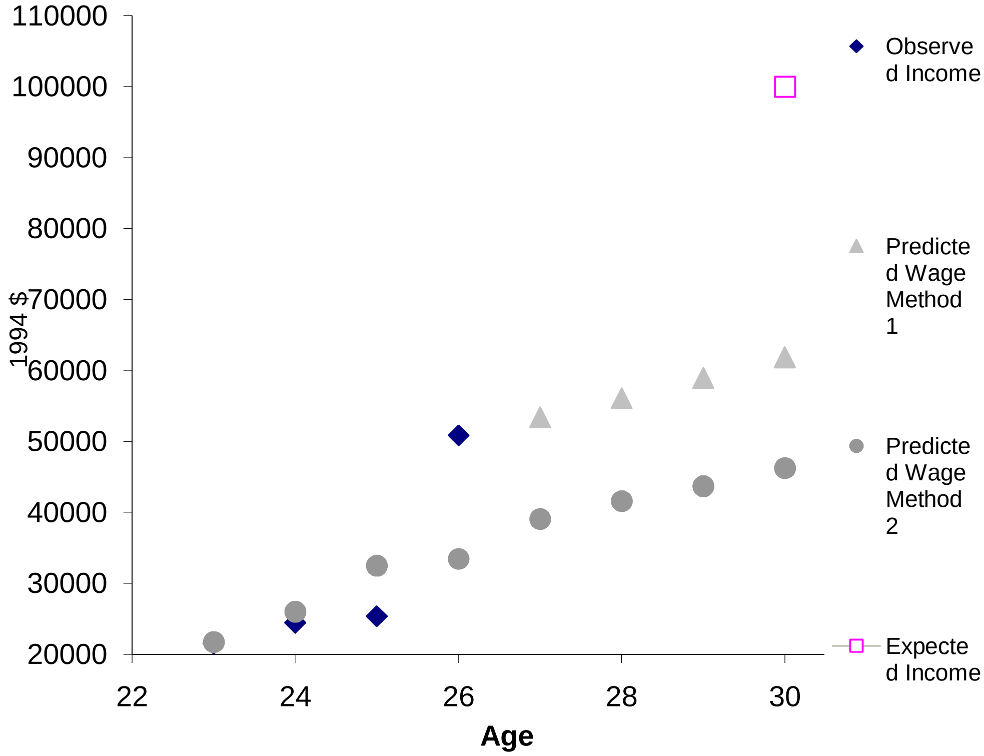
and

(b) there are a smaller proportion of respondents with below high school education

In a similar manner, I predict average wages for Hispanics to be around $27,000, while the CPS figure is closer to $20,000. This again could be due to slight differences in wording or response to the question regarding race and ethnicity in the CPS and NELS surveys. However, it is worth noting that both these groups only form a small part of the overall sample. The general message is that my prediction methods seem to generate a reasonable estimate of average age 30 wages.

In Figure A5, however, I show my two predictions for the illustrative NELS respondent. This highlights the difficulty of analysis at the individual level.

Figure A5. Comparison of wage prediction methods for ID 7286532 in the NELS



Notes:
See notes to Figure A3, A4 and A5

The two predictions of his age 30 wage are $15,000 (30%) apart. Given my discussion at the start of the section about whether the large jump in wages is temporary or permanent, anywhere between the two predictions, or even a figure outside of this range, could be possible. When dealing with group averages, over-estimates of wage growth for some individuals will be compensated by underestimates for others. However a comparison of expected and actual wages at the individual level is troublesome, as there is a large range of possible values for each individual's predicted age 30 wage.

Unearned Income

Thus far I have only considered wages; I now turn my attention to how other sources of finance contribute to total age 30 income. Some details were collected in the NELS about respondents' non-wage income at age 25 (the full previous year, 1999, prior to the survey). They were asked:

Without considering the earnings from employment that you just reported, approximately how much did you and your (spouse/partner) receive from other sources of income in 1999?
$...

These sources might include stocks and bonds, savings interest, insurance, alimony or child support, amily members, and disability payments

As this information is collected in just one question, measurement error may be a concern. Another difficulty is that the question asks for joint unearned income for the respondent and their partner (if married or cohabiting). Fortunately, the majority of those who did report a figure were not in a marriage or marriage like partnership. Appendix Table A9 presents the distribution of unearned income by marital status.

Table A9. Distribution of unearned income by marital status, for those reporting a value above zero

Percentiles of unearned income distribution	Single	Married/Cohabiting
1	30	100
5	200	200
10	500	500
25	1,200	1,100
50	4,000	3,250
Mean	8,528	7,335
75	10,000	7,000
90	20,000	15,000
95	30,000	30,000
99	60,000	65,000
Standard Deviation	16,045	13,298
% of observations where reported unearned income >0	25.7	19.0
% of observations missing	2.6	8.9
Observations (including 0's)	3,928	510

Notes:
Around 75% (80%) of single (married or cohabiting) individuals report 0 unearned income. The distributions above relate only to those who reported some form of unearned income (a value greater than 0)

The reported distributions of unearned income by married and single individuals are quite similar, though the former are more likely to not respond and the latter more likely to report 0.

The first question to ask is how much unearned income do young adults receive, and what proportion of total income does it make up at (a) the group and (b) the individual level? The second column of Table A10 provides details of the unearned income distribution at age 26 for NELS respondents. One striking feature is that the majority of individuals (74%) report no unearned income. It would appear that, even though unearned income is important conceptually, empirically it has relatively little influence on average total income (at least at the group level).

To investigate this proposition further, I again turn to the NLSY where, as opposed to the NELS, respondents are asked several questions about each aspect of their non-earned income. For instance, they were asked about their income from businesses, public support, educational grants and any other sources in a series of separate questions[63]. Table A10 compares the distributions of unearned income for NELS and NLSY sample members at age 26, for those reporting a value above 0.

Table A10. Distribution of unearned income at age 26 in the NELS and NLSY, for those reporting a value greater than 0

Percentiles of unearned income distribution	Reported unearned income in NELS	Reported unearned income in NLSY	Reported unearned income in NLSY for those reporting a value greater than $500
1	50	4	510
5	200	14	578
10	500	29	662
25	1,200	116	1,156
50	3,876	578	2,553
Mean	8,079	3,347	6,202
75	9,000	2,753	6,314
90	20,000	7,967	14,287
95	30,000	14,571	26,893
99	70,000	42,563	57,150
% reporting>0	26%	38%	20%

Notes:
All data are for individuals at age 26 in 1994 prices
Distribution is for respondents reporting a value greater than 0

[63] Note in the NLSY, respondents were asked separate questions about their spouses unearned income as well.

Notice firstly the NLSY has a greater proportion (38% compared to 26%) of people reporting positive unearned income. However, the distribution shows almost a quarter of these observations are less than $100. It seems that the NELS, by recording this data in a single question, misses many individuals who have a small quantity of unearned income. In any case, both the NELS and NLSY suggest that unearned income, *on average*, has only a small influence on total income at age 26. The median respondent indicates they have no unearned income. Even of the minority that do report a figure above 0, unearned income (on average) is relatively small compared to wages in most cases.

One may suggest that unearned income may make up a more significant proportion of total income at age 30 than at age 26. To investigate this, I compare mean wages to the mean total income for men in the 2003-2005 CPS March Annual Supplement[64]. On average (mean), unearned sources of finance contribute only $500 (2%) to total income. I performed a similar analysis on the NLSY 79 sample when they turned age 30, and found a similar result (unearned income makes a very small contribution to total income at the group or population level).

Overall it seems that, on average, unearned income makes up only a very small part of total age 30 incomes. Hence it should be of limited importance when one compares expectations to realisations at the group level. Therefore, to incorporate unearned income into my predictions, I simply use the value recorded at age 25 in the NELS. Implicitly this means that anyone with zero unearned income at 25 will also have zero predicted unearned income at age 30. Given its minor role, this should not introduce substantial bias at the group or population level.

[64] Several questions about other (unearned) sources of income were asked in the CPS. This includes how much they received from benefits, welfare, assistance, dividends and interest. The data I use is drawn from the CPS "Table Creator", available from http://www.census.gov/hhes/www/cpstc/cps_table_creator.html I produce two values, one looking at men's average wages, the other their total income. I assume that the difference between these two figures (average wages and average income) equals total income from unearned sources.

On the other hand, unearned income is a non-trivial matter at the *individual* level. Table A10 shows some individuals to report a figure over $10,000 at age 25 in the NELS data. But this could be a one-off inheritance from a relative dying, or sudden good luck with a stock option (especially given the technology boom at the time of the survey in 2000). There is no indication about how this unearned income may change in the future. Hence predicting unearned income at age 30 for a given individual is an even harder task than for wages. Thus the NELS simply does not contain the data to make estimation of unearned income at the individual level a realistic possibility.

Summary

Drawing together the results from this appendix, it seems that inferences at the group and population level should be reasonably robust to the problems identified with the NELS data. I have presented two methods to predict age 30 wages, which provide similar estimates of average wages at age 30, and that are comparable with external estimates from population level data. Moreover, even though age 30 unearned income is difficult to predict, I have shown that this makes up only a small proportion of total average income. I am therefore confident that the substantive inferences in section 5 regarding population and group level averages are robust to the data issues discussed throughout this section.

However, my concerns for analysis at the individual level remain. Figure A5 illustrates how two very different predictions, over $15,000 (30%) apart, can be made for any one individual. I have also assumed this person has no unearned income at age 30, as he did not report any at age 25. This would be quite a bold assumption to make. The implication is that inferences made at the individual level are likely to suffer from what may be quite severe biases. Consequently, I focus on group level analysis (mean and median outcomes), that I believe are robust to the assumptions I have made about the data. Though analysis at the individual level would be of great interest, I do not believe this to be sensible with the NELS data.

Appendix 3.

Comparison of OLS results using Method 1 to Method 2

Tables A11 to A13 provide regression results, analogous to these in Table 11, except that I now predict age 30 income using "Method 1". This method is described in more detail in section 4 and Appendix 2. Note that I implement Method 1 three ways, using different estimates of young adults wage growth from different surveys.

Comparing the results to those in Table 11, it seems that most of the patterns I describe in section 5 still hold. For example, specification 1 consistently shows that workers hold more realistic expectations than Art, Biology and Communication students. And, as described in section 5, the inclusion of the college drop-out dummies in specification 4 causes the subject coefficients to drop dramatically. However, it is worth noting that statistical significance has been lost for some groups of students in comparison to Table 11. For example, using the PSID to extrapolate wage growth ("Method 1b"), the coefficient estimates for Accounting, Finance and Biological Science groups are now only statistically significant at the 10% level (compared to the 5% level in Table 11). This seems to be a result of both a decrease in the estimated coefficient, and more variability in the data (recall my discussion of Table A7, where I show the standard deviation of predicted wages to be lower in Method 2 than Method 1). Nevertheless, I can confidently say that, on average, there is still little evidence that workers hold more realistic expectations of their future income than students. Moreover, although some of the coefficients have been reduced to lower levels of statistical significance, the general patterns found regarding specific groups of students still seem to hold. In particular, engineering, maths and computer science students hold more realistic expectations than workers (and those in creative disciplines) across all results. Likewise, I always find those who drop out of university have the least realistic expectations.

Turning to the other coefficients, there is again strong agreement across the prediction methods. Family income, and whether the student also holds a job at age 20, is rarely of statistical significance at any of the conventional levels. On the other hand, cognitive maths ability and the Black race dummy are always significant at the 5% level. Hence the general message from these tables is that the results presented in section 5 seem relatively robust to the prediction method that I use.

Table A11. OLS regression results comparing how realistic students are to workers (Prediction "Method 1a" using CPS wage growth estimates)

	Specification 1		Specification 2		Specification 3		Specification 4	
	Co	**SE**	**Co**	**SE**	**Co**	**SE**	**Co**	**SE**
Work-student status at age 20 (Ref: Working)								
Agriculture student	-0.10	0.08	-0.05	0.08	-0.07	0.07	-0.12	0.08
Economics, Finance student	-0.09	0.10	-0.02	0.11	-0.05	0.10	-0.12	0.10
Business, Management student	0.01	0.05	0.05	0.05	0.03	0.05	-0.05	0.05
Journalism, Communication student	0.20*	0.08	0.24*	0.08	0.21*	0.08	0.13	0.08
Computer Science, Maths student	-0.11	0.09	-0.06	0.09	-0.10	0.10	-0.20*	0.09
Education student	-0.02	0.06	0.02	0.06	0.00	0.07	-0.06	0.07
Engineering, Physical sciences student	-0.16*	0.05	-0.11*	0.05	-0.14*	0.05	-0.21*	0.06
Language student	-0.08	0.13	-0.02	0.13	-0.02	0.13	-0.10	0.14
Health student	0.19*	0.08	0.24*	0.08	0.22*	0.08	0.15*	0.08
Law student	0.44*	0.20	0.44*	0.19	0.44*	0.20	0.32	0.19
Biological science student	0.24*	0.08	0.30*	0.08	0.29*	0.08	0.21*	0.08
Social sciences, Humanities student	0.17*	0.07	0.22*	0.07	0.18*	0.07	0.10	0.08
Art student	0.31*	0.13	0.36*	0.13	0.35*	0.14	0.25*	0.13
Other student	0.05	0.06	0.08	0.06	0.06	0.06	-0.03	0.06
Not student or working	0.15*	0.06	0.16*	0.06	0.12*	0.05	0.12*	0.05
Missing	0.30*	0.07	0.33*	0.07	0.28*	0.07	0.30*	0.07
Maths ability at age 18	-	-	-0.07*	0.02	-0.06*	0.02	-0.05*	0.02
Race (Ref: White)								
American Indian or Alaska Native	-	-	-	-	0.07	0.08	0.06	0.08
Asian or Pacific Islander	-	-	-	-	0.06	0.07	0.06	0.07
Black, not Hispanic	-	-	-	-	0.22*	0.05	0.20*	0.05
Hispanic or Latino	-	-	-	-	0.11*	0.05	0.10*	0.05
More than one race	-	-	-	-	0.18*	0.07	0.17*	0.07
Missing	-	-	-	-	0.02	0.08	0.00	0.08
Family income parents reported when respondent was age 18 (Ref: Bottom quintile)								
2nd quintile	-	-	-	-	-0.03	0.05	-0.03	0.05
3rd quintile	-	-	-	-	-0.02	0.05	-0.01	0.05
4th quintile	-	-	-	-	0.08	0.05	0.08	0.05
Top quintile	-	-	-	-	0.03	0.06	0.05	0.06
Student at 20, who also held a part-time job (Ref: No)								
Yes	-	-	-	-	0.01	0.04	-0.02	0.04
College Dropout (ref: No)								
Yes	-	-	-	-	-	-	0.21*	0.05
Constant	0.44*	0.03	0.41*	0.03	0.38*	0.05	0.39*	0.05

Notes:
These results refer to when I use Rubenstein and Weiss (2007) CPS estimates of wage growth (see Table 6) to predict NELS sample members age 30 income.

Table A12. OLS regression results comparing how realistic students are to workers (Prediction "Method 1b" using PSID wage growth estimates)

	Specification 1		Specification 2		Specification 3		Specification 4	
	Co	**SE**	**Co**	**SE**	**Co**	**SE**	**Co**	**SE**
Work-student status at age 20 (Ref: Working)								
Agriculture student	-0.12	0.08	-0.07	0.08	-0.09	0.07	-0.15*	0.08
Economics, Finance student	-0.09	0.10	-0.03	0.11	-0.05	0.10	-0.14	0.10
Business, Management student	0.01	0.05	0.05	0.05	0.02	0.05	-0.06	0.06
Journalism, Communication student	0.20*	0.08	0.24*	0.08	0.21*	0.08	0.12	0.08
Computer Science, Maths student	-0.12	0.09	-0.06	0.09	-0.11	0.10	-0.22*	0.09
Education student	-0.03	0.06	0.01	0.07	-0.01	0.07	-0.08	0.07
Engineering, Physical sciences student	-0.17*	0.05	-0.11*	0.05	-0.15*	0.05	-0.22*	0.06
Language student	-0.08	0.13	-0.02	0.13	-0.02	0.13	-0.11	0.14
Health student	0.18*	0.08	0.23*	0.08	0.21*	0.08	0.13	0.08
Law student	0.44*	0.20	0.44*	0.19	0.44*	0.20	0.30	0.19
Biological science student	0.22*	0.08	0.28*	0.08	0.27*	0.09	0.19*	0.08
Social sciences, Humanities student	0.15*	0.07	0.21*	0.07	0.16*	0.07	0.07	0.08
Art student	0.30*	0.13	0.35*	0.13	0.34*	0.14	0.23	0.13
Other student	0.05	0.06	0.08	0.06	0.05	0.06	-0.05	0.06
Not student or working	0.14*	0.06	0.15*	0.06	0.12*	0.05	0.12*	0.05
Missing	0.31*	0.07	0.33*	0.07	0.28*	0.07	0.29*	0.07
Maths ability at age 18	-	-	-0.08*	0.02	-0.06*	0.02	-0.05*	0.02
Race (Ref: White)								
American Indian or Alaska Native	-	-	-	-	0.07	0.08	0.06	0.08
Asian or Pacific Islander	-	-	-	-	0.06	0.08	0.06	0.07
Black, not Hispanic	-	-	-	-	0.22*	0.05	0.21*	0.05
Hispanic or Latino	-	-	-	-	0.12*	0.05	0.10*	0.05
More than one race	-	-	-	-	0.16*	0.08	0.14	0.08
Missing	-	-	-	-	0.03	0.08	0.00	0.08
Family income parents reported when respondent was age 18 (Ref: Bottom quintile)								
2nd quintile	-	-	-	-	-0.03	0.05	-0.03	0.05
3rd quintile	-	-	-	-	-0.02	0.06	-0.01	0.06
4th quintile	-	-	-	-	0.08	0.05	0.08	0.05
Top quintile	-	-	-	-	0.02	0.06	0.04	0.06
Student at 20, who also held a part-time job (Ref: No)								
Yes	-	-	-	-	0.02	0.04	-0.02	0.04
College Dropout (ref: No)								
Yes	-	-	-	-	-	-	0.23*	0.05
Constant	0.43*	0.03	0.40*	0.03	0.37*	0.05	0.38*	0.05

Notes:
These results refer to when I use Rubenstein and Weiss (2007) PSID estimates of wage growth (see Table 6) to predict NELS sample members age 30 income.

Table A13. OLS regression results comparing how realistic students are to workers (Prediction "Method 1c" using NLSY wage growth estimates)

	Specification 1		Specification 2		Specification 3		Specification 4	
	Co	SE	Co	SE	Co	SE	Co	SE
Work-student status at age 20 (Ref: Working)								
Agriculture student	-0.15*	0.08	-0.09	0.08	-0.11	0.07	-0.18*	0.08
Economics, Finance student	-0.12	0.11	-0.05	0.11	-0.08	0.11	-0.17	0.10
Business, Management student	-0.02	0.05	0.03	0.05	-0.01	0.06	-0.10	0.06
Journalism, Communication student	0.18*	0.08	0.22*	0.08	0.19*	0.08	0.08	0.08
Computer Science, Maths student	-0.14	0.09	-0.08	0.09	-0.13	0.10	-0.25*	0.09
Education student	-0.06	0.06	-0.01	0.07	-0.03	0.07	-0.11	0.07
Engineering, Physical sciences student	-0.19*	0.05	-0.13*	0.05	-0.17*	0.06	-0.26*	0.06
Language student	-0.11	0.12	-0.04	0.12	-0.04	0.13	-0.15	0.14
Health student	0.16*	0.08	0.21*	0.08	0.19*	0.08	0.10	0.08
Law student	0.43*	0.20	0.43*	0.20	0.42*	0.20	0.26	0.19
Biological science student	0.20*	0.08	0.26*	0.09	0.25*	0.09	0.16*	0.08
Social sciences, Humanities student	0.13*	0.07	0.19*	0.07	0.14*	0.07	0.04	0.08
Art student	0.28*	0.13	0.34*	0.13	0.32*	0.14	0.20	0.13
Other student	0.03	0.06	0.06	0.06	0.03	0.06	-0.08	0.06
Not student or working	0.14*	0.06	0.15*	0.06	0.12*	0.05	0.12*	0.05
Missing	0.30*	0.07	0.32*	0.07	0.27*	0.07	0.28*	0.07
Maths ability at age 18	-	-	-0.08*	0.02	-0.07*	0.02	-0.05*	0.02
Race (Ref: White)								
American Indian or Alaska Native	-	-	-	-	0.08	0.08	0.06	0.08
Asian or Pacific Islander	-	-	-	-	0.05	0.08	0.06	0.07
Black, not Hispanic	-	-	-	-	0.22*	0.05	0.21*	0.05
Hispanic or Latino	-	-	-	-	0.12*	0.05	0.10*	0.05
More than one race	-	-	-	-	0.16*	0.08	0.15*	0.08
Missing	-	-	-	-	0.04	0.08	0.00	0.08
Family income parents reported when respondent was age 18 (Ref: Bottom quintile)								
2nd quintile	-	-	-	-	-0.03	0.05	-0.03	0.05
3rd quintile	-	-	-	-	-0.02	0.06	-0.01	0.06
4th quintile	-	-	-	-	0.07	0.05	0.08	0.05
Top quintile	-	-	-	-	0.01	0.06	0.03	0.06
Student at 20, who also held a part-time job (Ref: No)								
Yes	-	-	-	-	0.02	0.04	-0.02	0.04
College Dropout (ref: No)								
Yes	-	-	-	-	-	-	0.26*	0.05
Constant	0.42*	0.03	0.38*	0.03	0.36*	0.05	0.37*	0.05

Notes:
These results refer to when I use Rubenstein and Weiss (2007) NLSY estimates of wage growth (see Table 6) to predict NELS sample members age 30 income.