Can students predict their starting salary? Yes!

Webbink, D.; Hartog, J.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Can students predict their starting salary?
Yes!

Dinand Webbink*
Joop Hartog**

Abstract
We use Dutch panel data in which students have been asked to state their expected starting salary and confront these with realisations four years later. Both level and structure of expectations and realisations are remarkably close: we barely find systematic under- or overestimation effects

JEL code: J24; I21

* CPB Bureau of Economic Policy Analysis, The Hague, NL (h.d.webbink@cpb.nl)
** University of Amsterdam, NL (hartog@fee.uva.nl)

Comments by Bas Jacobs and Edwin Leuven are gratefully acknowledged.
1 Introduction

Expectations of earnings lie at the heart of the ‘human capital’ model. The model states that students, in deciding on the amount of education, compare the outcomes of the different options and choose the option with the highest return. Considering the pivotal role of expectations in the choice process, one would expect that the expectations of pupils and students have been studied extensively. Such is not the case, however. After the classic studies by Freeman (1971, 1975), only a few studies have been carried out in this field. Manski (1993) commented as follows: ‘The profession has traditionally been sceptical of subjective data; so much that we have generally been unwilling to collect data on expectations. Instead, the norm has been to make assumptions about expectations formation.’ Many economic researchers plainly assume equality of expectations and realisations (for example Willis and Rosen, 1979) without empirical testing.

Other disciplines in educational research like sociology or psychology have less problems with collecting subjective data in this field (see for instance Smith and Powel (1990), Berndt and Miller (1990), but also among economists this scepticism of subjective data seems to be fading, considering some recent studies (Dominitz and Manski (1994) and Betts (1996). One of the main features of these studies is that expectations of earnings are compared with realisations of earnings on the labour market by other individuals with similar characteristics. In this paper we compare individual expectations with realisations for the same individual, with data from the panel study 'Verder Studeren' ("Continued Education"). The analysis is relevant to test a common assumption in theoretical models and can also throw light on questions about over- or underinvestments in education. For instance, if students with certain characteristics have much higher realisations than expectations there might be under-investment in education.

As noted, studies on labour market expectations and the relation with educational decisions start with Freeman (1971, 1975). In these classic studies Freeman showed that the expectations of students correspond to a high degree with the performance of earlier cohorts on the labour market. This applies both to the initial wages for various occupations and to the wages after 15 years and at the end of the respective professional careers. The analysis also showed that expected income differences between occupations have an influence on the choice of education, assuming a limited set of educational alternatives.

Dominitz and Manski (1994) asked students in the U.S.A. to complete a computerised questionnaire in order to obtain information about income expectations for various levels of education. The main conclusions drawn from this explorative study were that students are capable of making realistic estimates of future incomes, and that the general expectation was that education leads to higher income. Betts (1996) analysed income expectations of undergraduates, and showed that there was no great divergence between expectations and realisations in the labour market (by others). Students in higher years proved much better informed with respect to the labour market than first-year students. In the Netherlands, the influence of income expectations on educational decisions was analysed in detail by Kodde and Ritzen (1986). However, they have not analysed the expectations themselves.

The economic literature on earnings expectations is small. Moreover, the validity of expectations has not been analysed at the individual level. In this paper we make a contribution to fill that gap. The next section discusses the data, section 3 gives the analysis and 4 concludes.
The data and frame of analysis

The data we use are from the longitudinal research project "Continued Education" ('Verder Studeren'). In particular, we use data from the panel among students in higher education started in 1991. It started with 3,845 students in higher education. The sample was stratified in two levels and nine types of higher education. With this sample all years, levels and types of higher education in the Netherlands were covered. In particular, it covers both university students and students in higher vocational education. Also, both freshmen and students in higher years were included. The students were questioned about their position and history in education and their motives for choosing this type of education. In the following years the sample was questioned every year about their position in or outside higher education, the motivation for the decisions made and their future plans. The last survey was held in 1995. During the project all students who participated in the first survey got all the subsequent questionnaires. This includes students who left the educational system or students who did not participate in one or more surveys. In each questionnaire several retrospective questions were asked. The answers to these questions made it possible to ‘repair’ gaps in the longitudinal data collection. Student participation was encouraged by offering prizes in each questionnaire. All these actions resulted in a modest panel-mortality even after five surveys. In 1995, the overall retention rate was 56%.

In 1991 the students were asked an open question about their expected starting salary after graduation. In 1995 more than 1,000 students (of the remaining 2,140 students in the panel) had left higher education and entered the labour market. These students were asked about their earnings. For 645 students we have information on both expected earnings in 1991 and realised earnings after graduation (for 355 students information is missing on at least one of the earnings variables). In the analysis of expected and realised earnings we only use variables known at the first survey in 1991. We distinguish four groups of variables:

- **background variables.** We have gender, age, parent's education, parental income;
- **higher education variables.** We have level and type of education, year of education, and part-time student status
- **secondary education variables.** We know secondary school marks, marks in languages science and humanities which are given on a 10-point scale. We also know whether the student repeated classes in secondary school and the schooladvice at the end of primary education
- **motivation in higher education.** We know study efforts in 1991, measured as weekly hours spent to class attendance, assignment, study, etc. We also have measures on extrinsic and intrinsic motivations, based on the psychological literature. The first measure is a weighted average of the scores on questions about the importance of labour market perspectives in choosing a study and is believed to proxy extrinsic motivation. The second measure is based on the answers to questions relating to interest in the contents of the study and is an indication of intrinsic motivation.

Students had also been asked, in 1991, their subjective probability to graduate, a variable shown to have good predictive performance for actual graduation (see Webbink, 1999, p 136).

Because not all students graduated at the same time or graduated at all we also used information on dropping out and work experience for calculating the income after graduation for students who started working well before 1995. The income in 1995 was corrected for the
returns on years of working experience and in case of drop out for the graduation premium (based on the regression effects of these two variables)

A full characterisation of the data is given in the Appendix. The data have been used for extensive analysis of educational choices in Webbink (1999) and in other publications of the research project (De Jong, 1997). Further details of data and related research findings are given in these publications.

3 Empirical analysis of expected and realised earnings

The first step of the analysis is the estimation of a model for the expected earnings of students in 1991 and for the realised earnings in 1995. The aim is to find out whether the structure of the determinants for expected earnings is the same as the structure for realised earnings. The estimation results of the OLS-regression are presented in Table 1. The dependent variables are the natural logarithm of expected and realised earnings.

Most of the effects on expected earnings have the same sign and comparable size as the effect on realised earnings. In fact, several coefficients are incredibly close. Female students expect to earn 6 percent less than male students do and this is realised four years later. Most remarkable is the coincidence of the effects of the types of education on expectations and on realisations of earnings. For many types of education this coincidence includes sign and size of the effects and the students’ expectations are remarkably precise. Students in economics, health and technical and law studies expect more income than students in social studies and earn more after four years. Students in languages and cultural studies might be disappointed with their earnings; they expect the same as students in social studies but earn considerably less. Students who repeat classes expect more and also realise higher incomes than other students. This positive effect on earnings is in line with the human capital view and not with the screening view as Oosterbeek (1992) points out.

The effects of marks in secondary education on expected earnings are not translated into realised earnings. Students with high scores in science subjects expect more income than others but do not earn more four years later. The opposite is true for students with high scores in humanity subjects: they do not expect more than others but realise higher incomes. We also see that students from high-income families expect higher incomes after graduation but do not earn more after four years than students from low-income families. Exposure to a high family income apparently generates too rosy expectations. It may be a factor in explaining higher participation in extended education for wealthy background students.

The main conclusion is that the pattern of effects on expected earnings is highly comparable with the effects on realised earnings. However, these effects don’t have to hold at the individual level. Therefore we compared expectations and realisations at the individual level in two ways. First by cross-tabulating the distribution of expected earnings and realised earnings (Table 2). Second, by analysing the systematic differences between expectations and realisations.

---

1 Predicting realized income from expected income has an adjusted R-square of 0.06 and a coefficient of 0.25 (t-value 6.7).
<table>
<thead>
<tr>
<th></th>
<th>expected earnings</th>
<th>realised earnings*</th>
<th>% difference**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coeff.</td>
<td>t-value</td>
<td>coeff.</td>
</tr>
<tr>
<td>intercept</td>
<td>6.824</td>
<td>23.25</td>
<td>7.268</td>
</tr>
<tr>
<td>female</td>
<td>-0.058</td>
<td>-2.43</td>
<td>-0.059</td>
</tr>
<tr>
<td>age</td>
<td>0.008</td>
<td>2.12</td>
<td>0.004</td>
</tr>
<tr>
<td>parents education</td>
<td>-0.015</td>
<td>-1.32</td>
<td>0.010</td>
</tr>
<tr>
<td>log parental income</td>
<td>0.052</td>
<td>1.98</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**higher education**

field of study (ref. social studies)

<table>
<thead>
<tr>
<th></th>
<th>coeff.</th>
<th>t-value</th>
<th>coeff.</th>
<th>t-value</th>
<th>coeff.</th>
<th>t-value</th>
</tr>
</thead>
</table>
economics                | 0.077  | 1.93    | 0.071  | 1.80    | 0.005  | 0.11    |
health/medical studies    | 0.112  | 2.68    | 0.112  | 2.73    | -0.004 | -0.08   |
agricultural studies      | 0.058  | 1.21    | 0.065  | 1.37    | -0.004 | -0.06   |
science studies           | -0.036 | -0.82   | -0.023 | -0.52   | -0.020 | -0.35   |
technical studies         | 0.075  | 1.67    | 0.089  | 2.02    | -0.019 | -0.33   |
languages/cultural studies| -0.027 | -0.52   | -0.125 | -2.41   | 0.098  | 1.44    |
educational studies       | -0.023 | -0.52   | 0.007  | 0.17    | -0.027 | -0.46   |
law studies               | 0.064  | 1.00    | 0.100  | 1.57    | -0.055 | -0.66   |
part-time study           | 0.109  | 2.63    | 0.022  | 0.55    | 0.097  | 1.82    |
weekly effort study hours '91 | 0.000  | 0.38    | -0.002 | -1.81   | 0.002  | 1.71    |

**motivation in '91**

<table>
<thead>
<tr>
<th></th>
<th>coeff.</th>
<th>t-value</th>
<th>coeff.</th>
<th>t-value</th>
<th>coeff.</th>
<th>t-value</th>
</tr>
</thead>
</table>
extrinsic motivation      | 0.015  | 2.34    | 0.008  | 1.37    | 0.005  | 0.06    |
intrinsic motivation      | 0.001  | 0.16    | -0.003 | -0.38   | 0.002  | -0.19   |
expected prob. graduation '91 | 0.001  | 0.64    | 0.002  | 2.37    | -0.001 | -1.18   |

**secondary education**

<table>
<thead>
<tr>
<th></th>
<th>coeff.</th>
<th>t-value</th>
<th>coeff.</th>
<th>t-value</th>
<th>coeff.</th>
<th>t-value</th>
</tr>
</thead>
</table>
average mark languages    | -0.022 | -1.30   | -0.016 | -0.97   | -0.006 | -0.28   |
average mark humanities   | 0.022  | 1.32    | 0.048  | 2.94    | -0.028 | -1.33   |
average mark science      | 0.030  | 2.16    | -0.007 | -0.50   | 0.038  | 2.08    |
school advice             | -0.005 | -0.75   | -0.001 | -0.15   | -0.004 | -0.49   |
repeated classes          | 0.069  | 2.91    | 0.055  | 2.39    | 0.017  | 0.56    |
educational position in '91 (ref. = university older)
| vocational freshmen      | -0.126 | -3.67   | -0.133 | -3.91   | 0.025  | -0.54   |
vocational older          | -0.155 | -5.40   | -0.115 | -4.07   | -0.037 | -1.01   |
university freshmen       | -0.014 | -0.33   | -0.008 | -0.18   | 0.021  | -0.38   |
years till graduation     |        |         |        |         | -0.032 | -2.47   |
adjusted r-square         | 0.13   | 0.12    | 0.01   |         |         |         |
# observations            | 645    | 645     | 645    |         |         |         |

* corrected for years of working experience and drop-out
** (exp-real)/exp
In Table 2 we see that the mean of expected income is about 10 percent higher than the mean of the realised income (mean of 7.77 versus 7.67) and that the standard deviations do not differ. This optimistic view of students was also found by Smith and Powel (1990). Each earnings distribution in Table 2 is divided in 6 classes (one, two, or more than two standard deviations below or above average earnings).

Most of the sample lies just below or just above the average. Only a small fraction of students (approximately 2%) lies in the right tail of expectations of earnings and the left tail of realised earnings or vice versa (shaded area in Table 2). This means that large differences between expectations and realisations are rare.

Table 2  
Expectations and realisations at the individual level (% of total sample, n=645)

<table>
<thead>
<tr>
<th>Expected earnings (lnW1) μ=7.77 σ=.29</th>
<th>realised earnings (lnW2) μ=7.67 σ=.28</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnW1 &lt; μ-2σ</td>
<td>lnW2 &lt; μ-2σ</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>μ-2σ&lt;lnW1 &lt; μ-σ</td>
<td>μ-σ&lt;lnW2 &lt; μ-σ</td>
</tr>
<tr>
<td>0.6</td>
<td>1.6</td>
</tr>
<tr>
<td>μ-σ&lt;lnW1 &lt; μ</td>
<td>μ-σ&lt;lnW2 &lt; μ</td>
</tr>
<tr>
<td>1.1</td>
<td>2.2</td>
</tr>
<tr>
<td>μ-σ&lt;lnW1 &lt; μ+σ</td>
<td>μ-σ&lt;lnW2 &lt; μ+σ</td>
</tr>
<tr>
<td>0.6</td>
<td>2.2</td>
</tr>
<tr>
<td>μ+σ&lt;lnW1 &lt; μ+2σ</td>
<td>μ+σ&lt;lnW2 &lt; μ+2σ</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>μ+2σ&lt;lnW1 &lt; μ+2σ</td>
<td>μ+2σ&lt;lnW2 &lt; μ+2σ</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

# observations 16 46 161 278 110 36

In the right hand column of Table 1 we analyse the systematic differences between expectations and realisations at the individual level. The dependent variable is the difference between ln expected earnings and ln realised earnings. Hence, testing the coefficients in this regression against zero on the difference of the coefficient for expected and for realized earnings. In the model we controlled for the time distance between the first survey and the moment of graduation because it is plausible that ‘years till graduation’ influence the validity of the earnings prediction by students (in the earlier regressions this effect is averaged in the intercepts). The main conclusion from this model is again that systematic differences between earnings expectations and realisations are absent. At the 5%-level we only find significant effect of average marks for science and years till graduation. Students with higher scores on science subjects overestimate their earnings after graduation. Students who come closer to graduation tend to become less optimistic about their earnings, reducing their expectation by some 2 1/2 % for every year closer to graduation.

The very poor overall fit of the model clearly shows that differences between expected earnings and realised earnings are not systematically related to variables commonly available, to type of study or to the additional, less common, variables that we have available.

The correlation between the systematic part and the non-systematic part of the earnings equations confirm these findings (Table 3).
Table 3  Correlation between systematic and stochastic part of expected and realised earnings

<table>
<thead>
<tr>
<th>expected earnings</th>
<th>realised earnings</th>
<th>systematic part</th>
<th>error term</th>
</tr>
</thead>
<tbody>
<tr>
<td>systematic part</td>
<td></td>
<td>0.81</td>
<td>0</td>
</tr>
<tr>
<td>error term</td>
<td></td>
<td>0</td>
<td>0.15</td>
</tr>
</tbody>
</table>

The correlation between the error terms of the expected earnings equation and the realised earnings equation is very low, indicating the absence of systematic differences between expectations and realisations.

4 Conclusions

At the individual level large differences between earnings expectations of students on their starting salaries and realisations by graduates are rare. Moreover, there are no systematic differences between expectations and realisations. These results show that findings by Dominitz and Manski (1994) that student are capable of making realistic estimates of future incomes not only hold at the group level but also at the individual level. Yes, students can make good predictions of their starting salaries several years before they actually graduate. Hence, their choices on continued education are well-informed choices.

A very interesting finding is that the structure of student's earnings expectations is very similar to the structure of realised incomes. Especially the coincidence of the effects of the type of education on earnings expectations and realisations is remarkable. Students almost perfectly anticipated that starting salaries are higher by 7% in economics, by 11% in health and medical studies, by 6% in agricultural studies by 9% in technical studies. Students in languages substantially overestimated their starting salary, students in law underestimated. Also, students from high-income families tend to overestimate their starting salary.

Generally we conclude that directly measured individual expectations of starting salaries quite accurately mimic the differences in starting salaries by type of study and by various individual characteristics.
references

Berndt, T. and K. Miller (1990), Expectancies, Values and Achievements in Junior High School, in: *Journal of Educational Psychology*, 82, pp. 319-26


Appendix

Table A1. Description of the data

<table>
<thead>
<tr>
<th>variable</th>
<th>description</th>
<th>mean</th>
<th>st. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>log expected earnings (ln W₁)</td>
<td>expected starting salary after graduation measured in October 1991, log of guilders per month net</td>
<td>7.77</td>
<td>0.29</td>
</tr>
<tr>
<td>log realised earnings (ln W₂)</td>
<td>realised earnings after graduation measured in December 1995, log of guilders per month net</td>
<td>7.67</td>
<td>0.28</td>
</tr>
<tr>
<td>female</td>
<td>female=1; male=0</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>age of student in October 1991</td>
<td>17.5</td>
<td>3.2</td>
</tr>
<tr>
<td>parents education</td>
<td>maximum educational level of father and mother; scale from 1 (primary education) to 5 (university education)</td>
<td>3.01</td>
<td>1.12</td>
</tr>
<tr>
<td>log parental income</td>
<td>guilders per month net of taxes and premiums</td>
<td>8.22</td>
<td>0.47</td>
</tr>
</tbody>
</table>

**Higher education**

field of study

<p>| social studies | yes=1; no=0 | 0.15 |
| economics      | yes=1; no=0 | 0.17 |
| health/medical studies | yes=1; no=0 | 0.13 |
| agricultural studies | yes=1; no=0 | 0.09 |
| science studies | yes=1; no=0 | 0.12 |
| technical studies | yes=1; no=0 | 0.14 |
| language/cultural studies | yes=1; no=0 | 0.06 |
| educational studies | yes=1; no=0 | 0.10 |
| law studies     | yes=1; no=0 | 0.04 |
| part-time study | yes=1; no=0 | 0.12 |
| weekly effort study hours '91         | average weekly hours spent on study                                         | 32.3  | 12.9     |
| vocational freshmen | yes=1; no=0 | 0.20 |
| vocational older       | yes=1; no=0 | 0.33 |
| university freshmen    | yes=1; no=0 | 0.09 |
| university older       | yes=1; no=0 | 0.38 |
| years till graduation  | years till graduation since October 1991                                  | 1.73  | 1.16     |</p>
<table>
<thead>
<tr>
<th>variable</th>
<th>description</th>
<th>mean</th>
<th>st. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>secondary education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>average mark languages</td>
<td>scale from 1 (lowest) to 10 (highest)</td>
<td>6.78</td>
<td>0.68</td>
</tr>
<tr>
<td>average mark humanities</td>
<td>scale from 1 (lowest) to 10 (highest)</td>
<td>6.94</td>
<td>0.72</td>
</tr>
<tr>
<td>average mark science</td>
<td>scale from 1 (lowest) to 10 (highest)</td>
<td>6.72</td>
<td>0.89</td>
</tr>
<tr>
<td>school advice</td>
<td>advice at the end of primary school; scale from 1 (lowest level) to 7 (highest level)</td>
<td>5.16</td>
<td>1.72</td>
</tr>
<tr>
<td>repeated classes</td>
<td>yes=1; no=0</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>motivation in '91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>extrinsic motivation</td>
<td>weighted average of scores on questions about importance of labor market perspectives in choosing a study</td>
<td>5.75</td>
<td>1.80</td>
</tr>
<tr>
<td>intrinsic motivation</td>
<td>weighted average of scores on questions about the importance of interest in the contents of the study on choosing a study</td>
<td>8.68</td>
<td>1.30</td>
</tr>
<tr>
<td>expected prob. graduation '91</td>
<td>expected probability of graduating in present study (indicated by student in 1991)</td>
<td>89.4</td>
<td>14.9</td>
</tr>
<tr>
<td># observations</td>
<td></td>
<td>645</td>
<td></td>
</tr>
</tbody>
</table>