Over-education and earnings: where are we, where should we go?

Joop Hartog

University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

Received 31 August 1997; accepted 12 October 1997

Abstract

Drawing on empirical studies from five countries spanning an interval of two decades, regularities in the incidence of over- and undereducation are outlined, as well as consequences for individual earnings. The results are confronted with three theoretical models (search, human capital and assignment), but none of these is convincingly related to the specification of the earnings function. Directions for further work are suggested. © 2000 Elsevier Science Ltd. All rights reserved.

JEL classification: J24; J31

Keywords: Overeducation; Earnings

1. Where are we?

The canonical Mincerian earnings function relates the log of earnings to years of schooling attained, experience and its square. Duncan and Hoffman (1981) have started a literature by distinguishing between an individual’s attained level of education and the education required in the job. From these two concepts they derived measures of over- and undereducation, and they estimated returns to these years of mismatch, as well as returns to required years of education. Other authors have followed in their footsteps and estimated comparable specifications. The approach is attractive for two reasons. The specification is straightforward, is easy to estimate and has a simple and direct interpretation. Second, the specification establishes a link with the demand side of the labour market (ignored in the canonical specification), with allocation processes and with immediately visible variations in rewards for individuals with identical schooling operating in different positions.

This paper will survey a number of issues relevant for that literature. In this section, it will consider conceptual issues and measurement problems and it will survey the empirical results that are available for the small set of countries that have been studied: the United States, The Netherlands, Portugal, Spain and the United Kingdom. It will discuss interpretations of the findings in Section 2. Section 3 focusses on the ignored issue of comparative static analysis. Finally, Section 4 concludes with a reflection on the possible routes for further research.

1.1. Conceptual issues

The common characterization of the demand side of the labour market, both in economic theory and in empirical applications, is quite crude. The standard specification of labour demand acknowledges the role of physical capital, often assumed homogenous, but has no other variables to characterize jobs (apart from wages and prices of course). Job amenities are taken into account in applied work on the theory of compensating

---

1 In this special issue of the journal, results are added for Canada and Germany.
differentials, but otherwise, the job is just some unspecified qualitative entity. Or, as Lazear (1995, p. 77) has put it: “The job has no role in traditional production theory”. Yet, jobs differ in many characteristics that affect productivity, and there is certainly information available in adjacent fields like psychology of occupations and personnel. A lot of work has been done on job requirements in job evaluation studies. There is a world waiting to be explored with data from personnel psychology. Job requirements cover many dimensions of worker abilities and aptitudes, and required schooling is only a single variable that necessarily compresses much information. Still, it’s a step in the right direction.

In the literature on overeducation (and undereducation), required schooling has been measured in three different ways:

1. from job analysis (JA): systematic evaluation by professional job analysts who specify the required level (and type) of education, for the job titles in an occupational classification. The most elaborate example is the United States Dictionary of Occupational Titles. The information was used by e.g. Thurow and Lucas (1972) and by Hartog (1980). The DOT information was used to estimate an overeducation earnings function by Rumberger (1987). Similar information for Portugal was used by Kiker and Santos (1991) and for the Netherlands by Oosterbeek and Webbink (1996). In the literature on overeducation (and undereducation), required schooling has been measured in three different ways:

2. from worker self-assessment (WA): the worker specifies the education required for the job. It may be a direct and explicit specification of the type of schooling required, or it may be indirect, stating whether, compared to the worker’s actual education, a higher or a lower (or a different) education is needed. This method has been applied by e.g. Duncan and Hoffman (1981), Hartog and Tsang (1987) and Sicherman (1991). There may be slight variations in the specification. Sicherman (1991) uses response to a question in the American PSID data: “How much formal education is required to get a job like yours?”. Alba-Ramirez (1993) uses the question: “What kind of education does a person need in order to perform your job?”. These are different definitions, although not necessarily perceived differently by respondents.

3. from realized matches (RM): required education is derived from what workers in the respondent’s job or occupation usually have attained, e.g. the mean or the mode of that distribution. This method has been applied by e.g. Verdugo and Verdugo (1989) and Groot and Maassen van den Brink (1995).

Conceptually, systematic job analysis JA is a very attractive source for defining job requirements, because of its explicit goal of objectivity, clear definitions and detailed measurement instructions (see for example the discussion in the Dictionary of Occupational Titles US Department of Labor, 1965). Even the scope for substitution of different levels and types of educations (and other traits) can be indicated. Job analysis really starts from the technology of the job, the type of activities to be performed. However, the required careful, systematic work may be too expensive to carry out on a large scale.

In fact, even subsequent editions of the DOT, published at wide time intervals, in many cases simply copy earlier analyses, and only a small share of the analyses is new. This makes assessment of changes in the job structure from information in the subsequent editions rather unreliable (Cain and Treiman, 1981). Translating the job requirements into a single schooling variable may bring substantial errors. Glebbeek (1993) has demonstrated for the Netherlands that assigning job level codes to survey responses on type of work has a large measurement error (a low correlation of repeated assignment). Verdugo and Verdugo (1992, p. 692) note that according to the DOT Handbook a single job analyst visits the job site and discusses requirements with the employer. They doubt both reliability and validity of the measure (compare also Cain and Treiman, 1981).

Worker self-assessment WA has the advantage of drawing on all local, up-to-date information. The assessment deals, in principle, precisely with the respondent’s job, not with any kind of aggregate. However, it is a measurement that lacks rigorous instructions. Individuals may easily overstate the requirements of their job, to inflate the status of their position. Or they may simply reproduce actual hiring standards. This causes problems if actual schooling levels in the labour force increase over time, and employers adjust hiring standards while the jobs themselves have not changed. Sicherman (1991) argues that such a bias is not the explanation for the overeducation that has been measured for the US. He compares the schooling of overeducated workers, 1976–1978, with the average education of their age-occupation cohort in 1970, and indeed finds that it is higher. And the similar comparison for undereducated workers finds that they have less education than their earlier counterparts: the effects are not due to a general rise in the education level (Sicherman, 1991, p. 104).

Using realized matches RM is far from the ambition to uncover the technological requirements of a job. It

---

2 For an elaboration, see Hartog (1994) and for an application with US data, Hartog (1980).

3 An early application of this method is Blaug et al. (1967, chapter VI), who relate to the literature on “manpower planning”: deriving the demand for workers by education from predictions of industrial and occupational structure.

4 A careful review of assessments of skill requirement changes in the US is given by Spenner (1985).
measures allocation, actual assignment practice as determined by hiring standards and labor market conditions. This allocation is endogenous, just as the quantity traded is endogenous in a standard textbook market model. Hence, if this information is used, it should be interpreted as the market result in an assignment model, not as something like a shift indicator of the demand curve. Realized matches should be interpreted quite differently from the other two measures, which do remind of demand curve location parameters.

Can we prefer one measure over the other? Usually, the choice is dictated by data availability: you use what is available. With full freedom to choose the research design, it would be best to opt for JA: let trained job analysts grade the jobs. They could focus on technology, and they should even be able to indicate scope for substitutions, between different educations or between schooling and on-the-job training or experience. But for large, up-to-date data sets with job specific information (as compared to aggregates such as occupations) this is a costly affair, and the cheaper method of worker self-assessment WA certainly has its virtues. Assuming required education as a single-valued variable is a meaningful variable to characterize a job (see below), the merits of available JA measurement, relative to WA measurement, depend on the level of aggregation, the time lag in the observation and the care and precision of the measurement procedure. There appears to be no reason to expect bias in any particular direction in JA measurement, except when analysts would be systematically drawn towards the actual schooling levels of workers observed on the job, or towards the hiring standards applied by employers: the actual allocations would then dominate the measures. WA measurement may be biased upward by workers’ inclination to inflate the standing of their job. It may be influenced by observed hiring standards for newly hired workers. With rising educational attainments over time, this would give an upward bias. Hence, there is some reason to fear upward bias in WA measurement.

Considering the relative merits of the three measures, JA is conceptually superior. But its actual measurement often does not meet the highest possible standard. It will generally be available only for specific years. Since it will be costly to implement, actual, up-to-date measurement is usually not available. WA will therefore often be the best available measure. RM contains observations on the equilibrium realized by the interplay of supply and demand: as a measure of the demand side it is inadequate.

1.2. What, typically, is the incidence of mismatch? How does it evolve over time?

Naturally, the incidence of mismatch depends on how we specify required schooling. Also, it will depend on the precise definition of over/undereducation. A common, and obvious, definition takes overeducation as actual schooling greater than required, with undereducation similarly defined. However, several authors follow Verdugo and Verdugo (1989) and require actual schooling to be more than one standard deviation (of the actual schooling distribution for the job/occupation) away from job requirements. This can only reduce the incidence of mismatch relative to the straight difference, as it builds in a large safety area of two standard deviations. Typically, such measures find symmetry: incidence of over- and undereducation are quite similar, at about 10 to 15%. Interestingly, this is just what one would find beyond the one standard deviation tails of a normal distribution. Thus, we may think of the Verdugo’s measure as picking out the tails of a normal distribution of realised matches.

Tables 1 and 2 collects results from various studies. It’s fairly common to find a proper match in about 60% of the cases, with a somewhat lower estimate for the United States and very low estimates for Portugal in some years. As noted, the RM method usually finds symmetry in under- and overeducation (Portugal 1991 is an exception). For the Netherlands, estimates from WA and JA can be compared for 1971, resp. 1974, and they are very similar. The difference is somewhat larger when we compare them with five years apart (1977–1982). For the United States, estimates can be compared for the same year, 1973, and now we find that the incidence of undereducation is estimated at the same percentage, while the difference for overeducation is very large. But in fact, the incidence of overeducation at 57%, estimated by the JA method, strikes as a very high rate. Also, the very low rates of proper matching in Portugal, for the JA measure, are reason for suspicion.

A comparison of individual evaluations by WA and JA has been made for the Netherlands by Van der Velden and Van Smoorenburg (1997). Their observations are for workers at one year after their graduation from secondary education or from higher vocational education. The correlation coefficient for the two estimates of required education is 0.66. If the WA measures are aggregated to occupation levels, to make them more comparable with the occupation based JA measures, the correlation coefficient increases to 0.78.

Although the measures of over- and undereducation are not very precise, the data invite to make comparisons over time. We should then only consider the WA and the JA measures, since the RM method, as a relative measure is not suitable for this purpose. We may conclude that for the three European countries that we have observations on, the incidence of overeducation has increased, that of undereducation has decreased. This holds for the Netherlands 1960–1995, for Spain 1985–

5 Santos (1992) and Groot (1997) also present comparisons.
Table 1
Incidence of over- and undereducation (percentages)

<table>
<thead>
<tr>
<th>Country</th>
<th>Netherlands</th>
<th>Spain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overeducation</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>Proper match</td>
<td>58</td>
<td>59</td>
</tr>
<tr>
<td>Undereducation</td>
<td>36</td>
<td>27</td>
</tr>
<tr>
<td>Evaluation method</td>
<td>JA</td>
<td>JA</td>
</tr>
</tbody>
</table>

Table 2
Incidence of over- and undereducation (percentages)

<table>
<thead>
<tr>
<th>Country</th>
<th>Portugal</th>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>1985</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overeducation</td>
<td>18</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Proper match</td>
<td>63</td>
<td>58</td>
<td>30</td>
</tr>
<tr>
<td>Undereducation</td>
<td>19</td>
<td>17</td>
<td>43</td>
</tr>
<tr>
<td>Evaluation method</td>
<td>RM</td>
<td>RM</td>
<td>JA</td>
</tr>
</tbody>
</table>

1.3. What, typically, are the earnings consequences of mismatch?

The overeducation literature has a standard earnings function which we may call the ORU specification, for Over-, Required and Undereducation. It is written as follows.

\[ \ln W = \gamma_0 S_o + \gamma_1 S_o + \gamma_2 S_o + \eta \] (1)

1990 and for Portugal 1982–1992. This suggests that the strong expansion of participation in education has outpaced the increase in the demanded levels of education. In the Netherlands, the observed trend will also have been influenced by the high exit rates of older workers with low levels of schooling. The development in the United States appears different. While overeducation may have followed a U-shaped pattern between 1969 and 1977, there is evidence that it decreased in the subsequent period, until 1984. Between 1978 and 1984, the extent of undereducation appears to have diminished.

\[ \text{This points to a trend opposite to that in the three European countries.} \]

\[ \text{7 The observations on the United States are confirmed in Daly et al. (2000).} \]
where \( S_r \) is the schooling years required in the job, \( S_o \) is the number of years of overschooling (individual’s attained education minus required education in the job if positive; otherwise, \( S_o \) is zero) and \( S_u \) is undereducation years (required minus attained schooling years if positive, zero otherwise). \( X \) contains other explanatory variables, \( \eta \) is a random error term. The standard Mincer equation is included as the special case where the \( \gamma' \)'s are identical: the decomposition of attained years of schooling is irrelevant. In some cases, \( \gamma_r \) and \( \gamma_u \) are constrained to be equal, in most applications this is tested statistically.

In Tables 3–5 we have collected results for ORU earnings functions from five countries, from different sources for different years, amounting to a set of 45 regression results. We have restricted the studies to those that measure under- and overeducation in years, thus eliminating the specification introduced by Verdugo and Verdugo (1989) (we will discuss their outcomes below). The survey of empirical studies supports the following conclusions.\(^8\)

1. The returns to required schooling are higher than the returns to actual education. This follows from a comparison between the standard Mincer specification with individual’s attained education and the ORU specification. The results cannot be read from the tables, but they hold for the United States (Duncan and Hoffman, 1981; Rumberger, 1987, “middle specification”, Hartog and Tsang, 1987; Sicherman, 1991), for Spain (Alba-Ramirez, 1993), for Portugal (Kiker and Santos, 1991), for the Netherlands (Hartog and Oosterbeek, 1988; Oosterbeek and Webbing, 1996) and for the United Kingdom (Groot and Maassen van den Brink, 1995; Sloane et al., 1995). The deviation is relatively large in the UK. Groot and Maassen van den Brink report 7.9 versus 4.5% for men and 9.4 versus 5.6% for women.

2. Returns to overeducation are positive, but smaller than to required education. Typically, the returns to overeducation are about half to two-thirds of the returns to required education. All reported coefficients are significantly different from zero.

3. Returns to undereducation are negative. The penalty for undereducation is always smaller than the returns to required education. It is usually also smaller than the returns to overeducation, but here exceptions occur. Moreover, the coefficient on undereducation is not always significantly different from zero. When tested statistically, the coefficients on over- and undereducation differ significantly.

4. These conclusions are not sensitive to the measure of required education. We have results for all three measures of required education: job analysis, worker assessment and realized matches. Explicit comparisons have only been made by Santos (1992, 1995) for Portugal and by Rumberger (1987) for the United States. The three conclusions given above hold independent of the type of measurement. Sometimes estimated returns are affected by the type of measurement, but sometimes they are virtually identical (Rumberger, 1987; Van der Velden and van Smorenburg, 1997). This conclusion is quite appealing in view of the measurement unreliabilities noted above.

Whenever the ORU specification was formally tested against the Mincer specification (all coefficients equal), the ORU specification passed that test (Duncan and Hoffman, 1981; Hartog and Oosterbeek, 1988; Alba-Ramirez, 1993; Sicherman, 1991; Sloane et al., 1995; Groot 1996, 1997; Vieira, 1997). The superiority of the ORU specification is not testimony of a non-linearity in the returns to (required) education, as several authors conclude from testing this. Santos et al. (1996, p. 23) found a positive term for education squared, with other results upheld. Alba-Ramirez (1993, p. 270) reports the same result: a positive coefficient for education squared, other results maintained. Sicherman (1991, p. 116) also reports that testing with education squared shows that education mismatch effects are not due to non-linearity in the returns to schooling. The ORU specification, if formally tested, also outperformed the “job competition” specification, with only required education relevant (Hartog and Oosterbeek, 1988; Groot 1996, 1997; Sloane et al., 1995; Vieira, 1997).

We can use the results to make two comparisons. First, in a given job, with a given level of required education, earnings for undereducated workers are lower and earnings for overeducated workers are higher, compared to workers who bring just the required education to the job. Second, workers with a given education who get a job requiring more education than they have, have higher earnings than those in a job just matching their education (overeducation has a positive return), but lower earnings than workers who have the required education for the job they share: overeducation has lower returns than required education.

Verdugo and Verdugo (1989) stand apart in this literature for two reasons. First, they use attained education as the benchmark in the earnings function, rather than required education. And second, for under- and overeducation they require a gap between attained education and the mean education of those in the same occupational classification (300 occupations) of at least one
Table 3  
The ORU earnings function

<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Gender: T,M,F.</th>
<th>Schooling required</th>
<th>Overeducation</th>
<th>Undereducation</th>
<th>N</th>
<th>Required education based on</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>T,M,F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WA</td>
</tr>
<tr>
<td>NL</td>
<td>1982</td>
<td>F</td>
<td>0.071 (15.4)</td>
<td>0.057 (8.1)</td>
<td>-0.025 (3.0)</td>
<td>540</td>
<td>WA</td>
</tr>
<tr>
<td></td>
<td>1982</td>
<td>F</td>
<td>0.076 (13.9)</td>
<td>0.065 (7.4)</td>
<td>-0.019 (1.9)</td>
<td>400</td>
<td>WA</td>
</tr>
<tr>
<td></td>
<td>1982</td>
<td>F</td>
<td>0.052 (5.9)</td>
<td>0.037 (3.2)</td>
<td>-0.040 (2.5)</td>
<td>140</td>
<td>WA</td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>?</td>
<td>0.061 (20.3)</td>
<td>0.028 (4.66)</td>
<td>-0.026 (8.66)</td>
<td>1654</td>
<td>WA</td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>M</td>
<td>0.092 (17.6)</td>
<td>0.052 (5.7)</td>
<td>-0.033 (4.1)</td>
<td>986</td>
<td>WA</td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>M</td>
<td>0.079 (10.4)</td>
<td>0.063 (5.3)</td>
<td>-0.011 (0.8)</td>
<td>740</td>
<td>WA</td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>M</td>
<td>0.092 (78.7)</td>
<td>0.040 (16.0)</td>
<td>-0.060 (30.2)</td>
<td>11597</td>
<td>WA</td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>F</td>
<td>0.033 (10.6)</td>
<td>0.019 (5.0)</td>
<td>-0.033 (8.3)</td>
<td>2046</td>
<td>WA</td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>F</td>
<td>0.136 (28.4)</td>
<td>0.012 (3.15)</td>
<td>-0.066 (9.73)</td>
<td>4686</td>
<td>WA</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>T</td>
<td>0.120 (18.6)</td>
<td>0.013 (2.52)</td>
<td>-0.059 (6.30)</td>
<td>2317</td>
<td>RM</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>M</td>
<td>0.0149 (22.7)</td>
<td>0.007 (1.31)</td>
<td>-0.083 (9.34)</td>
<td>2369</td>
<td>RM</td>
</tr>
</tbody>
</table>

Other variables included:
- A/E age/experience: E
- T tenure: T
- G gender: G
- Occ occupation dummies: Occ
- Tr training: Tr
- O other: O

Notes:
- a Hartog and Oosterbeek (1988), table 3
- b Groot (1997), table 3, column III, respondents' age between 43 and 63
- c Oosterbeek and Webbink (1996), table 1
- d Alba-Ramirez (1993), table 3, col. 2
- e Beneito et al. (1996), table 5
- f Courtesy of Dr. Wim Groot, University of Leiden and Amsterdam; data as in Groot and Maassen van den Brink (1995)
<table>
<thead>
<tr>
<th>Country</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
<th>Portugal a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender: T,M,F</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>Schooling required</td>
<td>0.060</td>
<td>0.060</td>
<td>0.086</td>
<td>0.083</td>
<td>0.081</td>
<td>0.098</td>
<td>0.101</td>
<td>0.072</td>
<td>0.054</td>
<td>0.073</td>
<td>0.038</td>
<td>0.085</td>
<td>0.077</td>
<td>0.107</td>
<td>0.108</td>
<td>0.114</td>
<td>0.080</td>
</tr>
<tr>
<td>Overeducation</td>
<td>0.053</td>
<td>0.044</td>
<td>0.063</td>
<td>0.054</td>
<td>0.023</td>
<td>0.019</td>
<td>0.038</td>
<td>0.029</td>
<td>0.039</td>
<td>0.031</td>
<td>0.041</td>
<td>0.027</td>
<td>0.041</td>
<td>0.047</td>
<td>0.073</td>
<td>0.089</td>
<td>0.087</td>
</tr>
<tr>
<td>Undereducation</td>
<td>-0.038</td>
<td>-0.034</td>
<td>-0.054</td>
<td>-0.052</td>
<td>-0.037</td>
<td>-0.051</td>
<td>-0.048</td>
<td>-0.056</td>
<td>-0.032</td>
<td>-0.024</td>
<td>-0.032</td>
<td>-0.014</td>
<td>-0.037</td>
<td>-0.035</td>
<td>-0.059</td>
<td>-0.076</td>
<td>-0.075</td>
</tr>
<tr>
<td>Required education based on Other variables included</td>
<td>RM</td>
<td>RM</td>
<td>RM</td>
<td>RM</td>
<td>JA</td>
<td>JA</td>
<td>JA</td>
<td>JA</td>
<td>JA</td>
<td>JA</td>
<td>JA</td>
<td>JA</td>
<td>JA</td>
<td>JA</td>
<td>JA</td>
<td>JA</td>
<td>JA</td>
</tr>
<tr>
<td>T tenure</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>Occ occupation dummies</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
<td>Occ</td>
</tr>
<tr>
<td>O other</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

a Santos (1995); 1985: Table 4.29 1991: Table 4.31. Regression equation includes interaction (over/under) education with tenure and experience; returns at sample means of tenure and experience. The direct school year coefficients significant at 1%, interaction coefficients sometimes not significant at 10%.

b Vieira (1997), (7-4-97); Other includes firm size and age, bargaining regime and 7 industry dummies

c Vieira (1997) (14-4-97)
Table 5
The ORU earnings function

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender: T,M,F.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schooling required</td>
<td>0.063</td>
<td>0.091</td>
<td>0.095</td>
<td>0.078</td>
<td>0.089</td>
<td>0.077</td>
<td>0.052</td>
<td>0.066</td>
<td>0.092</td>
<td>0.100</td>
<td>0.092</td>
<td>0.061</td>
<td>0.115</td>
<td>0.048</td>
<td>0.082</td>
</tr>
<tr>
<td>Overeducation</td>
<td>0.029</td>
<td>0.052</td>
<td>0.062</td>
<td>0.044</td>
<td>0.051</td>
<td>0.054</td>
<td>0.031</td>
<td>0.043</td>
<td>0.068</td>
<td>0.057</td>
<td>0.059</td>
<td>0.028</td>
<td>0.061</td>
<td>0.039</td>
<td>0.059</td>
</tr>
<tr>
<td>N</td>
<td>2034</td>
<td>1187</td>
<td>1327</td>
<td>1306</td>
<td>1311</td>
<td>669</td>
<td>368</td>
<td>669</td>
<td>3133</td>
<td>3898</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Required education based on</td>
<td>WA</td>
<td>WA</td>
<td>WA</td>
<td>WA</td>
<td>WA</td>
<td>WA</td>
<td>WA</td>
<td>WA</td>
<td>WA</td>
<td>WA</td>
<td>JA</td>
<td>JA</td>
<td>WA</td>
<td>WA</td>
<td></td>
</tr>
<tr>
<td>Other variables included</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/E age/experience</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>T tenure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G gender</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occ occupation dummies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tr training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O other</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

^a Duncan and Hoffman (1981)
^b Hartog and Tsang (1987)
^c Rumberger (1987); WA regressions: Table 6. JA regressions: Table 3. All coefficients are significant at 5%; under- and overeducation constrained to equal coefficients,
^d Sicherman (1991), Table A1, column c
^e Cohn and Khan (1995), Table 2, column 3
standard deviation. In their own empirical work they single out over- and undereducation cases with dummies, rather than measuring the gap in years. They find a negative effect for overeducation and a positive effect for undereducation. Santos (1995) uses the same specification for Portugal with data for 1985 and finds the same result. Gill and Solberg (1992) have shown that these results are not at variance with the standard results. Because conditioning is on attained education, the negative effect of the overeducation dummy simply brings out that overeducated workers are in a lower level job than those without overeducation. Of course, the magnitude of the coefficient cannot be compared with the coefficient on attained years of education: it is unclear to how many years of overeducation the dummy coefficient refers. The Verdugo and Verdugo dummy specification is less attractive. Building up a body of comparable research, over time and place, would be facilitated by deleting this specification.

It would obviously be highly informative if we knew the effect of over- and undereducation on productivity, rather than on wages. Several studies have established a negative effect of overeducation on job satisfaction (Tsang et al., 1991). Tsang (1987) has estimated that reductions in aggregated job satisfaction reduce plant output, thus implying that overeducation reduces worker productivity. According to this estimate, work reorganisation that raises job requirements would increase output by 8% for a year reduction in each worker’s overeducation. It is regrettable that this is the only study to focus on productivity effects. It leaves the puzzle that overeducation should reduce wages rather than increase them.

2. How should we understand the evidence?

There are three candidates for an interpretation of the findings: searching for the good match in an environment of imperfect information, a human capital interpretation and the hedonic/assignment interpretation. The interpretations are not mutually exclusive. “Mismatch” can be a temporary status in worker’s career development, resulting from imperfect information and adjusted by deliberate search. Human capital theory has two arguments to accommodate the observations: deliberate investment and the effects of omitted human capital components. A more extensive discussion of the three interpretations and full reference to the empirical sources is given in Hartog (1997). 11

2.1. Searching and matching

There is indeed evidence that mismatch is related to an individual’s experience in the labour market, and in that sense a temporary situation. From simple tabulations, or mean values of variables, in cross-section data, and from multivariate analyses it is clear that the incidence of overeducation falls with increasing age and experience, and that the incidence of undereducation at the same time decreases. The effect of job tenure however, is not unambiguous and there is certainly no unequivocal support for tenure to have the same monotonic effect as experience. Jointly, these results support the interpretation of mismatch as originating in imperfect information coupled with search for higher job levels during an individual’s career.

Suppose that workers with a given level of education seek to improve the required education level of their job. Given the positive effect of required education on wages this is equivalent to wage maximizing mobility behaviour. In wage functions, estimating the effects of tenure and experience is marred by potential biases from endogenous mobility. That literature can teach us on the effects to be expected with respect to over- and undereducation (Garen, 1988; Altonji and Williams, 1997; Teulings and Hartog, 1998, chapter 6). With voluntary job mobility, experience will bring about an increase in job level. Individuals only move if their job level increases, and in fact their actual job level at any time will be the highest they have been able to obtain; choice behaviour rules out a reduction. The case is different for job tenure however. With random job level offers, independent of job level already obtained, there are two opposing forces, one leading to a positive effect of tenure on job level (required education), the other leading to a negative

9 Groot (1996), with data for the United Kingdom in 1991, uses required education as a benchmark, defines the incidence of over- and undereducation by a gap of at least one standard deviation and then measures over- and undereducation as the difference in years between actual education and required education (the mode in the respondent’s occupation). He then finds negative effects for overeducation and positive effects for undereducation, contrary to all results in the literature. Note that his method imposes a discrete jump for earnings at the tails of the over/undereducation distribution for given required distribution. At my request, Groot kindly estimated a “standard” ORU specification on his 1991 data for the UK. As reported in Tables 3–5, this produced a “standard” result: sign reversals are typical for the Verdugo threshold measure. The result suggests the relevance of a test for non-linearities in the returns to over-and undereducation.

10 Cohn (1992) draws the same conclusion.

11 Overeducation has also been linked to sorting models, in which education is used as information on an individual’s expected capabilities (and productivity) that cannot readily be observed. It is not immediately clear how wage effects of under- and overeducation fit in. At a first glance, one would expect the wage to be related to attained education only, certainly for labor force entrants. On the sorting model, see Weiss (1995).
effect. If the present job has resulted from a particularly high offer, the worker will remain in this job. High job offers survive, workers move on from low offers, and the high tenures are characterised by the high job levels. The opposing force is the high outside offer: workers observed with short tenures are those who have just moved due to a high outside offer taking them away from their previous job. Now, low tenures are related to high job levels. Without specific distributional assumptions, it is not clear which factor dominates, and whether the effect of tenure will be positive or negative. The empirical findings thus fit in with the random search view. Job levels are indeed found to increase with experience, while the effect of tenure has been found to be positive, negative or absent. It would be even more convincing to be able to explain the variation in findings on tenure effects from differences in the underlying process, in different search and mobility cost etcetera, but this could only be speculation. It requires more information than the available studies provide. Surely, the imperfect information-search story goes a long way in explaining the observations we have, but we should note that the search story is consistent with many views of the labour market and does not settle any dispute.

2.2. Human capital

In the human capital perspective, overeducation may result from a deliberate choice because the low-level job is a good investment opportunity. Sicherman (1991) tests the prediction that overschooled workers are more likely to move to higher level occupations (measured from wage regression coefficient weighted sums of schooling, experience and required training). The prediction is borne out, but the same result of increased likelihood of upward occupational mobility is found for underschooled workers. Alba-Ramirez (1993) finds that overeducated workers are more likely to change occupation when moving firm and that undereducated workers are more likely to move to a different job in the same occupation. He explains this from the choice for investment in general human capital in the former case, and specific human capital in the latter. But search processes, or temporary excess supply in a market, could also explain this. With some education type in excess supply, young graduates could easily start with overeducation in their own line, and then decide to move to a different occupation. More specific evidence is needed to demonstrate superior explanatory power of this hypothesis. One such piece of evidence is given by Kiker et al. (2000): the overeducated have faster earnings growth with tenure.

Incomplete measure of the worker’s human capital has strong a priori support. Experience and training may indeed be thought to compensate for “undereducation”.

But as noted above, a positive effect of experience on job level (and a negative effect on overeducation) may also result from search under incomplete information. And with training, the problem is that the empirical studies often do not measure actual training, but training as required for the job, irrespective whether the worker got it or not. Required training in fact points to the job level, and more required training for undereducated workers is almost tautological: if you are undereducated, you are in a high level job, which requires more training. Beneito et al. (1996) use training as “once engaged in specific training for the job held”, and this indeed would seem to be the proper specification. Years undereducated in the present job significantly increase the probability of having had training specific to the resent job, years overeducated significantly reduce it. Groot (1997) also uses a measure of actual training: “did you receive on-the-job training at the current firm?”. This training measure significantly increases the likelihood of being undereducated, but has no effect on the likelihood of overeducation. If the training was for the present job, it is likely that the causality was from match quality to training, but it is also possible that the training was given for an earlier job and that from there, the worker has been promoted into undereducation.

There is also a bit of evidence that other worker qualities are related to under- and overeducation. Hartog and Jonker (1996) find that undereducation at age 43 is weakly positively related to childhood IQ, although at age 53 the effect has become insignificant. Hartog et al. (1996) find that both for men and women, the likelihood of underschooling significantly increases with quantitative literacy, while the likelihood of overschooling is not significantly affected. Thus, within their group, the underschooled are the more able (controlling for education and experience). Quantitative literacy is measured as the skill in dealing with information, a measure developed for the IALS project (OECD/Statistics Canada, 1995).

2.3. Assignment

The ORU specification of the earnings function reminds of the assignment literature (Sattinger, 1993), and indeed every presentation includes the reference. The reference easily comes to mind because assignment theories are very explicit about the relevance of the demand side of the labour market. Measuring match quality is in line with attention for the assignment of heterogeneous workers to heterogeneous jobs. And the theory allows for allocations that leave “tension” between the worker’s level of attributes and the level required in the job, even in equilibrium (Tinbergen,
1956). But the suggestive kinship does not immediately carry over in formal relations.

In the Tinbergen (1956) model wages are only related to the requirements of the job, with parameters determined by the underlying distributions of supply and demand. Worker attributes such as attained education are absent, and a fortiori the difference between actual and required education. In the more general specification of the hedonic wage equation there is a similar problem. Any characteristic traded in the labour market is rewarded at a point of tangency of an iso-profit function and an iso-utility function (a bid-price function and a demand-price function), the supply and demand distributions determine the selection of equilibrium points and the wage gradient along the realised values of a characteristic reflects the common slope of the iso-functions of the partners matched at that particular point. A fully specified model should identify both the wage function, the underlying supply and demand functions and the distributions of the supply and demand parameters. No one has estimated such a model yet (for estimates of partial models, see Van Ophem et al., 1993). In this model, again there is no place for the difference between attained and required education. Attained and required education can both be included. Individuals have indifference curves for wages and required education of the job and these curves may shift with the individuals’s attained education. Similarly, firms have iso-profit curves for wages and workers with different levels of attained education and these curves may shift with required education of the job. The shifts of the curves will affect the equilibrium wage envelope, passing through tangencies of iso-profit curves and iso-utility curves. But the difference between required and attained education has no particular meaning.

Recently, Teulings (1995) has presented a new specification for the wage function in an assignment model. We may call it the model with the “zipper allocation”. Workers are ranked top-down by quality (skill), jobs are ranked top-down by quality (complexity), and the equilibrium allocation zips them top-down together (cf Sattinger, 1975). The best worker goes to the most complex job, and the zipper slides down to match the two sides, until at least one side is exhausted (generating underemployed labour or vacant jobs at the bottom). The wage function just follows along and is a gradient on the teeth of the zipper. In the empirical specification, variables characterising the worker are taken together in an index of worker skill, variables characterising the job are taken together in an index of job complexity. With a single index on both sides of the labour market, the allocation problem has been reduced to a single dimensional matching problem. The wage function can be related to the demand side index or to the supply side index, but not to both at the same time: the two indexes map one-to-one into each other. Attained education can be included in the supply side index of skill, required education in the demand side index of complexity, but they should not both be related to the wage rate in a single equation, and the difference has again no particular meaning.

Within the general hedonic model, it is possible to develop an interpretation of the ORU earnings function, but ironically, that should either be a demand side interpretation or a supply side interpretation. Let a job be characterized by fixed required education. Individuals with different levels of education can be assigned to it. The individual’s education is taken as fixed, so there is no iso-utility function for supply of education. For a given job level, the reward to attained education traces out the value of education to the employer; the iso-profit curve. If, in a log specification, this effect is identical across all jobs, and the location of this curve is shifted with the level of required education, the ORU earnings function is an iso-profit function restricted to a common shape across jobs. We would expect the curve to be concave: the penalty on undereducation should be larger than the reward for overeducation. This holds indeed for the United Kingdom and it holds for Portugal if occupational dummies are included. It does not hold for the Netherlands, the United States and for Portugal without occupational dummies. Note that in this case there is a specific form of comparative advantage. The returns to education depend on the job, and hence, the earnings difference between individuals with different education is not constant: it depends on their assignment. Comparative advantage takes the restricted form of constant relative differences.

One can mirror this case to the supply side, by indexing individuals with their level of education, taking job levels as fixed and letting individuals select their optimal job level. An earnings function then traces out the individuals’s common iso-utility function for job level, shifted by their level of education. The reference level should now be attained education rather than required education. If individuals like underutilisation and dislike overutilisation, an excess in required education should increase wages, a shortage should reduce wages, as compensating variation for effort levels. This is precisely what Groot (1993) finds. But the function is not convex, as we would expect for an iso-utility function.

The upshot of this discussion is quite clear. The kinship between the ORU model and the equilibrium assignment models is obvious. But the ORU wage function is no immediate representation of these models. Connecting the two requires special assumptions and this limits the domain of applicability.

3. An extension: comparative static analysis

Published work sofar has separately studied the incidence and the earnings effect of overeducation, but
changes over time have not been considered. Yet with large changes taking place in the composition of the labor force and presumably of labor demand, it would be interesting to observe and understand the effect of such changes. Estimates from the ORU earnings function can be compared over time for the Netherlands, Portugal and the United States; Spain cannot be included, because the estimated earnings functions differ too much in the controls that have been included (adding occupational dummies picks up some of the effects of returns to education). In each of these three countries, returns to required education increase. For US men, between ’76/’77 and ’84, we see that the slopes for under- and overeducation suggest the opposite lands, is quite remarkable.

rather than a decrease, both for Portugal and the Netherlands confirm that including occupation dummies reduces all coefficients in the US is underestimated in Tables 3–5, for all controls for occupation. However, the results for Spain and Portugal confirm that including occupation dummies reduces all education coefficients. Hence, the increase in education coefficients in the US is underestimated in Tables 3–5, for all three coefficients.

that the observed changes are identical to the predictions. These results make clear that the situation in the United States is different from that in Europe, a conclusion that is also relevant for the discussions on the changes in the rate of return to attained education that have been going on for some time now. We will return to this issue below.

With the data for Portugal, we can also perform a straightforward test of the Tinbergen model. This model specifies explicit equations for comparative static analysis, admittedly in a very simplified structure, however. For the single variable case applied here it can be summarised as follows. Jobs are characterised by required education. Workers evaluate potential jobs by a utility function linear in log earnings and the squared deviation between actual and required education: both over- and undereducation reduce their utility, and have to be compensated in earnings. Workers choose a job (required education) to maximise their utility. The earnings function serves to solve the allocation problem: a normal distribution of required education is to be matched with a normal distribution of attained education. It is quadratic in required education of the job, with the parameters determined by the parameters of the two frequency distributions: means and standard deviations. To be specific:

\[ W = \lambda_0 + \lambda_1 S + \frac{1}{2} \lambda_2 S^2 \]

with \( S \), as before, indicating years of education required in the job. The parameters are given by

\[ \lambda_1 = \alpha \left( \frac{\sigma_a}{\sigma_i} \right) \]

(3)

\[ \lambda_2 = \alpha \left( 1 - \frac{\sigma_a}{\sigma_i} \right) \]

(4)

where \( \sigma \) is the rate of marginal substitution between earnings and match quality, specified in the utility function, and \( \mu \) and \( \sigma \) are mean and standard deviation for available (a) and required (r) education. The resulting job assignments are given by

\[ E_i = \mu + \sigma_i \left( E_{ai} - \mu \right) \]

(5)

where \( E_{ai} \) is the required education in the job that individual \( i \), with education \( E_{ai} \), will hold in equilibrium.

Now, for Portugal we have data on almost the entire labor force (from all private sector firms with more than one employee)\(^{16}\), with observations on individual edu-

\(^{13}\) An analysis of changes over time in the United states is given by Daly et al. (2000). The contribution by Groot and Maassen van den Brink (2000) also looks at changes over time.

\(^{14}\) The two earnings functions differ because only one uses controls for occupation. However, the results for Spain and Portugal confirm that including occupation dummies reduces all education coefficients. Hence, the increase in education coefficients in the US is underestimated in Tables 3–5, for all three coefficients.

\(^{15}\) The results on Portugal are presented in greater detail in Hartog (1997). I am grateful to Jose Vieira for the calculations. Further analysis is given in his Ph.D. thesis (Vieira, 1999).

\(^{16}\) The same dataset, for a single year, is used by Kiker et al. (2000).
Table 6
Testing the general excess supply-excess demand hypothesis

Europe: excess supply

<table>
<thead>
<tr>
<th></th>
<th>predicted</th>
<th>82–95&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Portugal 85–91&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Portugal 82–92&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
</tr>
<tr>
<td>required education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>overeducation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>undereducation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RM</td>
<td>RM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>JA</td>
<td>JA</td>
</tr>
</tbody>
</table>

US: excess demand

<table>
<thead>
<tr>
<th></th>
<th>predicted</th>
<th>76–84&lt;sup&gt;a&lt;/sup&gt;</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>required education</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>overeducation</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>undereducation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Results from Tables 3–5
<sup>b</sup> Results from Tables 3–5: Santos (1995)
<sup>c</sup> Results from Tables 3–5: Vieira (1997) (7-04-97) and (14-4)

cation and education required for their job, for three years. We can thus calculate means and the standard deviations for the two distributions, for 1982, 1986 and 1992. Assuming $W$ constant, we can then predict from Eqs. (3) and (4) what should happen to the coefficients in earnings Eq. (2), i.e. we can predict the signs of the changes over time, and we can compare this to the results from actually estimating Eq. (2) by OLS. We can do the same thing for job assignment: predict the direction of change of the coefficients in Eq. (5) from directly observing means and standard deviations, and compare this with changes in actually estimated coefficients (an OLS regression of required education on actual education). The results are summarised in Tables 7 and 8. For the earnings function, for the total sample, the changes as predicted from the underlying distribution parameters are borne out in three out of six cases. For men, two out of six changes in the coefficients of the earnings function have the right sign, for women the score is 3 out of 6. For the allocation function, the signs of the changes agree in two out of three cases if we apply the model separately for men and for women. If we take them together, the signs of the changes only agree in one of the three comparisons. Hence, not a very impressive score. Probably, the test can be made more sophisticated, but that would require additional modelling, which is certainly not straightforward.

The simple exercise conveys an important suggestion. To understand changes in the structure of allocation and earnings, we should not restrict our analyses to effects of changes in mean level of attained education and mean level of required education. While the Tinbergen specifications may be too restrictive, it certainly makes sense to look at dispersion on both sides of the labour market. In Portugal, in 1982, the mean level of required education was higher than the mean level supplied by the labour force. In the decade that followed, the increase in the mean supplied level was greater than the increase in the mean required level. But the rate of return to actual education has increased, from 5.2% to 6.4%, instead of decreased (Vieira, 1997). The explanation may indeed lie in the development of dispersions. Initially, supply and demand were equally dispersed, but in 1992, demand was more dispersed than supply. A wider dispersed demand distribution has apparently pulled up the returns. In 1982, the return to required education was 10.7%, in 1992 it was 13.2%. Hence, we should understand the increased return to required education from the increase in the demand dispersion relative to the supply dispersion, not from changes in mean levels. Stated otherwise:

Table 7
Earnings function<sup>a</sup>

<table>
<thead>
<tr>
<th></th>
<th>predictions $\sigma/\sigma_\mu$, $\mu_\mu$, $1-\sigma/\sigma_\mu$</th>
<th>estimates $\lambda_1$, $\lambda_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma/\sigma_\mu$, $\mu_\mu$, $1-\sigma/\sigma_\mu$</td>
<td>$\lambda_1$, $\lambda_2$</td>
</tr>
<tr>
<td>men</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1982–1986</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>1986–1992</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1982–1992</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>women</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1982–1986</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>1986–1992</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>1982–1992</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1982–1986</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>1986–1992</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>1982–1992</td>
<td>–</td>
<td>+</td>
</tr>
</tbody>
</table>

<sup>a</sup> Hartog (1997)


Table 8
Assignment function

<table>
<thead>
<tr>
<th></th>
<th>men</th>
<th></th>
<th>women</th>
<th></th>
<th>total</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma/\sigma_e$</td>
<td>estimated slope</td>
<td>$\sigma/\sigma_e$</td>
<td>estimated slope</td>
<td>$\sigma/\sigma_e$</td>
<td>estimated slope</td>
</tr>
<tr>
<td>1982–1986</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>1986–1992</td>
<td>-</td>
<td></td>
<td>+</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>1982–1992</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

Hartog (1997)

the human capital model in its purest form, with homogeneous human capital (i.e. perfect substitutability between different schooling levels) may work for the United States but it does not work for Portugal.

So, what is the story on Portugal emerging from these analyses? Over the decade following 1982, overeducation in Portugal increased. But since the return to education increased rather than decreased, we cannot take this as evidence of a homogenous expansion of human capital towards excess supply. If we look at the relative developments of dispersion on the demand side and the supply side, we see that demand dispersion outperformed supply dispersion. It is this extension of the demand distribution relative to the supply distribution that can explain the increase in the return to required education.\(^\text{17}\)

Decomposition analysis shows that this is a substantial contribution to the increase in wage dispersion that took place. We can decompose the contribution to changes in log wage inequality over time on the basis of the ORU earnings specification, as given by Eq. (1). The decomposition includes the regression weighted variances of required, over- and undereducation, experience (squared) and covariance terms. The change in the dispersion of required education, weighted by the earnings coefficient, explains 3/4 of the change in earnings inequality between 1982 and 1992 for men and 90% for women. The next largest contribution is given by the change in the residual variance.

4. Where should we go?

The ORU earnings function has proven itself as an extension of the canonical Mincerian earnings function, by passing statistical testing in several countries for several datasets and several years. The Mincer earnings function is generally accepted as a useful summary expression of the returns to education, a good starting point for comparisons over time and place. The ORU specification can serve the same purpose, of summarizing the effect of education on earnings, and pointing to interesting and relevant differences across labour markets. The robustness of the four regularities stated above supports this use.

In terms of proper estimation of the ORU earnings function, there is still a lot of interesting work to be done. The standard Mincer earnings function is now heavily under scrutiny as subject to bias when estimated with OLS. Measurement error, omitted variables and endogeneity of schooling can each have this effect (Card, 1995). For the ORU earnings function, in particular the first two sources are relevant. Required education will be measured with substantial error; there may be differences between the type of measurement (WA versus JA) that can be studied. It would also be interesting to have direct assessments of changes in job requirements in given jobs due to technological change. While it may not be easy to find good instruments for required education, it is perhaps easier to instrument under- and overeducation, for example from data on individual abilities and motivation. The modestly available evidence on the relevance of individual abilities for the incidence of undereducation indeed supports this.\(^\text{18}\) It is certainly important to allow for endogeneity of the worker-job matching and to estimate a simultaneous model for under/overeducation and earnings. This makes the ORU territory a very interesting playing field for econometric exercises. Some have already begun (Groot, 1997), but much more can be done.\(^\text{19}\)

Overlooking the findings from the ORU research, an interesting hypothesis can be tentatively formulated. It seems that some asymmetry is shaping up. Undereducation is less severely punished than overeducation is rewarded. Some studies find that mobility is more stimu-

\(^\text{17}\) Kiker et al. (2000), use Portuguese data for one year and conclude that skill-biased technological change explains observations on overeducation and earnings. This also stresses the dominance of demand side factors.

\(^\text{18}\) Evidence by Vella and Gregory (1996) suggests something similar: overachievers on schooling (more education than predicted) end up in the low end of the earnings distribution for their education group.

\(^\text{19}\) Individuals’ attained schooling may also be better measured than as former schooling years. Drop-out years, measured as years in school not topped off with a diploma may have separate impact. See Hartog (1983).
lated by undereducation than by overeducation. Among the undereducated, we find the more able individuals, while the overeducated are not characterized by a deviant distribution in terms of ability and skill. The undereducated have more training, the overeducated have not less. This suggests a process in which the workers arrive, searching, at some job of a particular job level. Then, in a sort of creaming process, the abler workers move up, and they may float upwards to a job level that makes them undereducated. But the other individuals that remain enjoy some protection: they are not kicked out of their job, their salary is not reduced. It’s like a model where there is some insurance of a minimum level based on initially assessed quality, where the best move up, but the less able are not systematically pushed down. It’s a process that also suggests a relation to the model of tournaments. Perhaps, being undereducated can be viewed as identifying the winner in a tournament; if so, the premium for the undereducated relative to those with the same schooling level who are not undereducated represents the prize of the tournament. This would open up possibilities of testing tournament theory.

Compared to the Mincer specification, a strong feature of the ORU specification is the explicit attention for the demand side variables. In conventional earnings functions, the demand side is completely ignored, except for the occasional dummy for occupation and industry. At the very least, job requirements provide a better basis for an index of demand shifts. In supply-and-demand analysis, such as that by Katz and Murphy (1992) and by Blau and Kahn (1996), shifts in demand for education are measured as a fixed-weight shift in industries, and some occupations. Katz and Murphy use 50 industries and 3 occupations, Blau and Kahn 6 industries and also no more than 3 occupations. Demand indexes based on job requirements are much richer. Job based indexes are more adequate than industry-based indexes, as demand for education will vary much more by job than by industry, and information is widely available, and is even country-specific. That prevents the need for such drastic assumptions as those made by Blau and Kahn.

The most important problem of ORU analysis is the lack of a coherent theoretical framework. It’s incidence has been interpreted as a search process emanating from imperfect information and as deliberate investment in the good opportunities of matches implying overschooling. The search interpretation certainly has strong credentials. However, there is no well-developed general equilibrium framework for interpretation of the earnings function. The ORU earnings function does not fit in a straight human capital model, as this would predict only attained education to be rewarded. Required education can only play a role if heterogeneity of human capital is admitted. But then it is not clear how this heterogeneity relates to the required education of the job.

ORU analysis is in the spirit of assignment models. But the ORU function does not directly derive from any of these models. One can design a set of assumptions that provide a connection. One may then attempt to use this for understanding changes over time, in a comparative static analysis. But this is not straightforward. Under appropriate assumptions, the ORU earnings function can be interpreted as an iso-profit function. One may loosely sketch a theoretical framework in this vein. In a single-dimensional framework, workers are allocated top-down, as in the zipper allocation. The ORU earnings function will reflect earnings at the equilibrium assignments. But it is not immediately clear how changes in the distributions of supply and demand translate into changes in the coefficients of the earnings function. The situation is even more complex if other variables than education are included in the analysis.

The lack of an operational theoretical framework is manifest in particular when it comes to analysing changes over time. The analysis for Portugal has shown that considering the changes in the dispersion of attained and required education can help to understand changes in earnings inequality. The increased inequality in Portugal can be understood from the demand dispersion increasing faster than the supply dispersion, thus elongating the earnings dispersion. This is an interpretation that is quite close to the hypothesis that the demand for higher education has expanded more rapidly than supply. Strong relative demand expansion is the common hypothesis used for explaining increased inequality in several OECD countries (Freeman and Katz, 1995). Perhaps, the two explanations may be viewed as a continuous and a discrete version of the same story. Hence, it would be interesting to confront the two approaches in greater detail.

4.1. So, where should we go?

For empirical research, it seems that the next step should be to concentrate on studying changes over time, and look at the simultaneous development of the incidence of over- and undereducation, characterising changes in the supply and demand distributions and estimating ORU earnings functions for different points in time. There is also a need for more refined econometric work, in particular to account for measurement error.

In terms of theory there is a need for extensive reflection on the interpretation of the results. As it stands, the ORU earnings function only fits in with assignment theory under very specific assumptions. It would be most useful to develop a framework for comparative static analysis. Only with a good understanding of the connection to more structural models of the labour market will there be a future for the ORU analyses. Only then will it be possible to fully exploit its advantage over the simple supply-and-demand model, with its strong assumptions.
necessary to cover-up the lack of direct measurement of changes in demand.

Acknowledgements

This paper is based on my Keynote speech for the Conference of the Applied Econometrics Association, Maastricht May 15–16 1997. I am grateful for the inputs, as comments, discussions or in even more substantial format, of Jose Vieira, Jesse Levin, Nicole Jonker, Dennis Jonker, Wim Groot, Hessel Oosterbeek and Jules Theeuwes.

References


