SCHOOLING, EARNINGS AND RISK

A research agenda and some references

joop hartog

Universiteit van Amsterdam

SCHOLAR, TINBERGEN INSTITUTE, AIAS

Abstract

In this paper, we reflect on the impact of uncertainty on standard issues in the theory of human capital: demand for education and returns to education. We also reflect on implications for an altogether ignored issue, the optimal design of school systems or curricula. The conclusions are summed up as a research agenda.

Part of this note was written while I was a Visiting Scholar at the Stanford Graduate School of Business, March 2002. It served as a background for lectures on Human Capital in Mannheim Summer School, May 15-17 2002.

first draft, 19-02-99; this version, 11-07-02

FILE: RISKTHEORYstanford2

Preliminary and provisional; not for quotation
1. INTRODUCTION

The labour economics literature is remarkably lopsided in dealing with issues of uncertainty. Uncertainty is extensively studied in cases of asymmetric information as the quest for an optimal employment contract. The second post-war revolution in labour economics entails a reworking and seeking a new understanding of the employment contract, focussing on observable institutional features that before were simply ignored. The internal labour has been discovered as a very interesting object of research. Most other interactions between agents are indeed first scrutinised for their structure of information available to agents. But the combination with the first revolution, the introduction of human capital theory, is seldom made. Yet, individuals embarking on an education face a host of uncertainties that have barely been analysed. It’s the purpose of this note to lay out important issues, draw on the small literature that is available and outline which problems are worthy of further research and how one might proceed.

An individual deciding on education is confronted with a series of uncertainties. First, the individual will have very imperfect information on the education, or alternative educations. She does not know what the programmes exactly entail, whether she will like them or not, and whether she will be able to fulfil the requirements. She may lack the ability or perseverance to complete any given programme. Second, after graduating from some programme, she faces a similar uncertainty in the labour market. Even being educated for some trade or profession, she may lack the ability or other requirements for success. Or, stated more generally, she does not know where in the distribution of occupational competence and earnings she will end up. Moreover, the market value of given occupational skills may fluctuate, with the business cycle, or in structural patterns of innovation and decline. Thus, in the terminology of Levhari and Weiss (1974), there is large “input uncertainty” (what are my abilities, what’s my input in the human capital production function) and large output uncertainty (what is the reward to my inputs).

Given these evident basic uncertainties, it is quite remarkable how few analyses actually deal with them. We can outline several important issues that deserve attention. Such an outline precisely coincides with the outline of this note.

1. How to deal with uncertainty in general?
2. Can we measure individuals’ attitude towards risk?
3. What risk is involved in choosing an education?
4. How does this risk affect the choice of an education?
5. Does the labour market respond to the investment risk?
6. Are there any implications for designing school structures and curricula?

2. HOW TO DEAL WITH UNCERTAINTY?

There are now three main models to deal with decisions under uncertainty: expected utility, rank dependent expected utility and the related cumulative prospect theory, and the direct Mean-Variance approach (cf. Powell, Schubert and Gysler, 2001).
2.1 Expected utility

The most widely used approach in applied work on uncertainty is the expected utility model EU, where alternatives are evaluated by the sum of the probability weighted utilities of the possible outcomes. The utility function is concave and the extent of concavity is used as a measure of individual risk attitudes, as defined by Arrow and Pratt. In the strict version of the theory, the probabilities are objective probabilities.

Expected utility theory has come under attack because it appears at variance with choices that people make when confronted with hypothetical alternatives in a laboratory or simple survey setting, and that are easily verified to match one’s intuition. Kahneman and Tversky (1979) give a convincing and easily accessible account of such cases (including the so-called Allais Paradoxes).

EU may miss the essence of risk appreciation. Suppose, an individual has no access to the capital market. She faces several options. In one, she has a fixed income of 100 in the next period and the period thereafter. In the other, period one yields 90 and period 2 yields 110. With the discount rate at zero, both options have a present value of 200. In utility terms however, with a concave utility function, the present value of utility is lower in the fluctuating income case than in the fixed income case, due to declining marginal utility. Now, suppose the second option involves an equal probability of 90 and 110 in both periods. The expected present value will again be 200. As before, the options are equivalent in monetary terms. Evaluated in utility terms the fluctuating and the risky option both differ from the fixed income option, but the utility evaluations of the fluctuating and the risky option are identical. In both cases, the present value of expected utility equals U(90) + U(110). The example can easily be generalised. A known but fluctuating income is evaluated in exactly the same way as an unpredictable, risky income prospect. This does not seem to catch the essence of risk. Wakker (1994) elaborates formally on this distinction between individual’s response to the probabilities of an alternative and the valuation of pay-offs (declining marginal utility).

2.2 Rank dependent expected utility

Rank dependent expected utility (RDEU) was developed by Quiggin (1982), to overcome the Allais paradoxes. Generally, the model starts out from:

\[
V(Y) = \sum_{i=1}^{n} w_i(p)U(x_i)
\]

\[
\sum_{i=1}^{n} w_i(p) = 1
\]

And where \( p \) is the vector of all \( p_i \), thus making \( w_i \) a function of all \( p_i \). The weights are then more particularly specified as

\[
w_i(p) = f\left(\sum_{j=1}^{i} p_j\right) - f\left(\sum_{j=i+1}^{n} p_j\right)
\]
Where the outcomes $x_i$ have been ranked such that $U(x_1) > U(x_2) > \ldots > U(x_n)$. The decision weights are defined on the marginal benefit of the event: the difference between an outcome at least as good as $x(i)$ and strictly better than $x(i)$.

Lattimore, Baker and Witte (1992) present a specification amenable to empirical testing. Their evaluation function is

$$V(Y) = \sum_{i=1}^n w(p_i) U(x_i)$$

where $U(x_i) = x_i^\alpha$

and

$$w(p_i) = \frac{\alpha p_i^\beta}{\alpha p_i^\beta + \sum_{k=1}^n p_k^\beta}$$

The evaluations are applied separately to gains and to losses. Clearly, with $\alpha = \beta = 1$, we are back to EU. The value of $\beta$ determines an inflection point: With $\beta < 1$, $w(p) > p$ up to the point $p_1^*$, where $w(p_1^*) = p_1^*$, and the reverse for $\beta < 1$. $\alpha = 1$ implies a "certain" weight function, with weights adding up to unity, $\alpha < 1$ and $\beta = 1$ implies "subcertainty", where weights add up to less than unity. In the latter case, stochastic dominance may be violated, as the individual may prefer $10$ for sure to a .50 chance each of winning $10$ or $10$, and he may even prefer $10$ for sure to a .50 chance each of winning $10 + e_1$ or $10 + e_2$, both $e_1$ and $e_2$ small.

In an experimental setting, one may confront participants with given specifications of $p$ and $x$ and then ask for their certainty equivalent. With sufficient number of observations per individual, one may then estimate the parameters for individuals. Lattimore, Baker and Witte do this for a gain specification (choosing between two burglary targets, one with a probabilistic coffer content, one with a content to be specified as the certainty equivalent, and one for a loss function with certainty equivalent in terms of maximum days of imprisonment to plead guilty). Probability transformation is common, some two thirds of the individuals have both $\alpha$ and $\beta$ smaller than one.

2.3 (Cumulative) Prospect theory

In Prospect theory (PT), developed by Kahneman and Tversky (1979), outcomes are not evaluated in terms of final wealth, as in EU, but in terms of losses and gains relative to a reference point. The utilities of the possible outcomes are weighted by their objective probability; this is later generalised (Tversky and Kahneman, 1992) to decision weights in terms of cumulative probabilities, as in RDEU. The utility function is concave for gains and convex for losses. Formally, in Cumulative Prospect Theory (CPT), we can write

$$V(Y) = \sum_{i=1}^n w^g_i u(x_i^g) + \sum_{i=n_y+1}^n w^l_i u(x_i^l)$$

(4b) with $$w^g_i = w^g \left( \sum_{n=1}^i p_n \right) - w^g \left( \sum_{n=1}^{i-1} p_n \right)$$
where the g stands for gains and the index l for losses.

The utilities are defined for outcomes relative to the reference point zero:

$$u(x_i) = \begin{cases} x_i^n & \text{if } x_i \geq 0 \\ -\lambda (-x_i^b) & \text{if } x_i < 0 \end{cases}$$

Hence, the utility functions for losses and gains have different slopes. $\lambda$ defines the relative slope of the two functions under and above the reference point. $\lambda > 1$ implies a steeper function for the loss section below the reference point than for the gain section above the reference point, which is said to reflect loss aversion.

With data obtained from an experiment, with two alternatives, they estimate the decision weights for the specification

$$w(p) = \frac{p^\gamma}{\left(p^\gamma + (1-p)^\gamma\right)^{1/\gamma}}$$

### 2.4 Risk/Variance models

Whereas the EU model derives risk aversion, a dislike for uncertainty, from the concavity of the utility function, and hence from decreasing marginal utility of wealth, the Risk/Variance or Mean/Variance or Mean/Standard deviation model (MV), developed by Markowitz, specifies a utility function directly in terms of mean and risk of the returns. Risk is commonly specified as the variance or the standard deviation. Hence,

$$V(Y) = u(M,V)$$

with the first and second derivative to M positive and negative, and the first derivative to V negative. The second order derivative to V might be positive, but one may allow for different patterns of increasing and decreasing risk aversion. Instead of the statistical, objective, variance one may also use a subjective measure of perceived risk.

Under conditions, MV and EU analyses are equivalent. In particular, if the underlying random variables have a normal distribution, MV and EU representations of preferences are equivalent, as was demonstrated by Meyer (1987), see Appendix. For an application, see Jacobs.

### 2.5 Which model?

In applied work on choice under uncertainty, one has thus several models available. The EU model is still widely used, and has the virtue of linking up with a large body of literature. But
clearly also, it has painful weaknesses, and perhaps in the end is not a convincing model to
deal with risk. The Allais Paradoxes show that individuals’ choice behaviour easily violates
EU theory. EU does not seem to catch the essence of risk; the equivalence of situations of
risky and of certain but fluctuating income shows that EU leans heavily on declining
marginal utility of wealth, which is barely the essence of risk. With the valuations linear in
probabilities, differences in individual risk attitudes can only originate in different curvature
of their utility function, which is not very attractive: one would expect differences in
appreciation of the probability distributions to be also relevant. In view of the appeal of the
alternatives and their apparent amenability to empirical applications, it seems actually quite
remarkable that the alternatives are not more widely employed.\footnote{CPT is applied by Donkers et al (1999) to analyse survey response questions on risk attitudes and by Beetsma
and Schotman (2001) to analyse choice behaviour of participants in a television game.} The generalisations of
RDEU and CPT are then easily at an advantage, with their increased flexibility from a non-
linear weighting of the probabilities of possible outcomes. However, Wakker (1994) notes
that there exists no convincing empirical support for RDEU, and also points to a weakness of
the specification. A change in outcome x does not affect the decision weights if it does not
upset the rankings, but if it does, the decisions weights are suddenly affected.

In cases of continuous rather than discrete probability distributions the MV approach seems
an attractive candidate (although Tversky and Kahneman (1992) point out that CPT can be
generalised to continuous distributions). Its equivalence to EU, under specified conditions,
seems a disadvantage, however, as it then would share its weaknesses as well. One may of
course generalise, add a measure of skewness to the utility function, ignore possible
equivalences, argue that risk and skewness are primitives in individuals’ preferences, and
specify an attractive utility function in mean, standard deviation and skewness.

3. CAN WE MEASURE ATTITUDES TOWARDS RISK?

Under EU, risk attitudes are defined by the curvature of the utility function. Risk aversion is
commonly defined by the Arrow-Pratt measure, based on the second derivative of the utility
function. Similarly defining “skewness affection”, or prudence as it is called in the finance
and lifecycle consumption analysis, based on the third derivative, is straightforward (Hartog
and Vijverberg, 2001). The parameters of these attitudes can be estimated from revealed
preference in a structural model of choice under uncertainty (lifetime consumption behaviour,
participation in lotteries and horse race betting, see Hartog and Vijverberg for references).
They can also be measured directly, from soliciting reservation prices for specified lotteries
in surveys (Hartog et al, 2002 and references cited there). The challenge with direct
measurement, of course, would be to demonstrate predictive validity for actual choice
behaviour, as in Cramer et al (2000).

Under RDEU and CPT, the parameters of the decision weight function and the utility
function, such as in (3)-(4) and (6) can be estimated from experimental (survey) data, as the
examples above indicate. They might in principle also be retrieved from properly specified
and estimated structural models. Remarkably, applications of the latter approach are very
rare.

Direct and revealed preference measurement of risk attitudes has established some empirical
regularities, but no narrowly confined parameter estimates. In the financial literature, the
observed long-term equity premium over riskless assets, requires a high coefficient of relative risk aversion to be consistent with individual choice theory based on EU, a value of at least 10. Such high values are not commonly found in structural revealed preference estimates. They are more likely to be in the 2 to 3 range, although exceptions occur. Direct measures, from survey questions on reservation prices of risky choices, tend to find higher values (see Hartog and Vijverberg, 2001, for references). When tested, utility functions with constant relative risk aversion, are usually not supported by the data (and constant absolute risk aversion can be ruled out on a priori grounds). Perhaps one of the strongest regularities is higher risk aversion for women than for men, or, more carefully: “it is a well-known stereotype that when confronted with risky decision situations, women choose low-risk alternative whereas men choose alternatives with higher risks” (Powell, Schubert and Gysler, 2001, who analyse different sources for this empirical regularity).

Experimental work on RDEU and CPT generally finds loss aversion (the utility function in losses is steeper than the utility function in gains) and decision weights that differ systematically from objective probabilities

With questionnaire methods, yielding sufficient variation, one may test the CPT evaluation of risk against EU. Similarly, with observations on perceived risk for a set of options from which individuals make an actual choice (such as continue in school or not), one may test the CPT model against EU for explaining individuals’ choice, as the EU model is nested in the CPT model. Interestingly, there is a growing interest in direct measurement of risk attitudes, e.g. plans to include a lottery reservation price question in the NLS project (Curtis Eberwein, private communication).

4. WHAT RISK IN CHOOSING AN EDUCATION?

As noted in the introduction, uncertainty when choosing an education abounds. The individual does not know what the education entails, whether she will like the program or not, has the ability to fulfil the requirements, whether she will like the occupation or set of occupations that is associated with the education, what the returns to the education will be when she enters the labour market and how they may fluctuate over the period that spans her future career. Hence, it will be easy to lay out a long list of imperfectly predictable variables and associated issues to deal with when considering educational choice. But that would not be very productive. Rather, we should focus on outlining relevant and manageable key issues.

When considering to enrol in an education, the individual faces at least three prime uncertainties. As to the education under consideration, she does not know whether she will complete the education (drop-out risk) and she does not know where in the post-school earnings distribution she will end up. As to the alternative of not going to school, she does not know for sure either where in the associated earnings distribution she will be located. These uncertainties complicate individuals’ choices in real life, but they also complicate the researcher’s life, as he has to adjust the model to analyse these choices.

4.1 drop-out risk

The consequences of drop-out risk for the rate of return can be easily demonstrated in some simple formulas, derived under varying assumptions. In each case, we assume full long-term
equilibrium, where the wage for the educated adjusts until lifetime (expected) earnings are equal.

*standard Mincer*

Assume, unschooled receive $Y_o$ per year during their entire working life, while after $s$ years of schooling the annual wage is $Y_s$. With discount rate (or cost of capital) $\delta$, the mark-off $M_s$ in $Y_o = (1 - M_s) Y_s$ can be solved as

(8) \[ M_s = 1 - e^{-\delta s} \]

This is the standard Mincer result. Now assume that at the end of $s$ years of schooling, graduates are separated from those who fail, with constant future annual earnings $Y_s$ and $Y_f$, and a probability $p$ to pass. The same reasoning then yields

(9) \[ M_s = 1 - e^{-\delta s} + (1 - p) \frac{Y_f - Y_s}{Y_s} e^{-\delta s} \]

This formula illustrates Adam Smith's fifth condition for differences in wages: the probability of success. As he states so eloquently: "In a perfectly fair lottery, those who draw the prizes ought to gain all that's lost by those who draw the blanks." (see section 6.1 of this paper). As the formula shows, the rate of return contains compensation for earnings postponement, and compensation for the expected earnings loss when failing rather than passing the final exam. An implication is that the rate of return for those who succeed should not only be estimated as a coefficient on schooling $s$ but also on the expected earnings loss from failure: probability multiplied by relative earnings loss.

*expected utility maximization*

If we allow for valuation in terms of the utility of earnings $U(Y)$, rather than straight earnings maximization, a similar development leads to

(10) \[ M_s = (1 - e^{-\delta s}) \frac{1}{e^u} + (1 - p) \frac{Y_s - Y_f}{Y_s} e^{-\delta s} \]

The first term reflects the move from earnings maximization to utility maximization, by dividing by the income elasticity of utility. With constant relative risk aversion, this elasticity would be a constant. The compensation for failure is the same as above because we use a first-order Taylor expansion to derive the result. As before, considering only the first term for those who succeeded in fact implies setting $p=1$.

*applying CPT*

If we apply CPT, the reference point to assess gains and losses, in the case of risky schooling, would be $Y_f$: schooling generates $Y_f$ with certainty. With probability $p$, there is a gain of $Y_s - Y_f$. Using the common specification $U(Y) = Y^\gamma$, the valuation of schooling will be
We can solve this, by equating the value of schooling and no-schooling as

\[ V(Y) = \frac{\alpha p^\beta}{\alpha p^\beta + (1-p)^\beta} (Y_s - Y_f)^\gamma + \frac{\alpha (1-p)^\beta}{\alpha (1-p)^\beta + p^\beta} (Y_f - Y_f)^\gamma \]

which can also be written as

\[ \frac{Y_s - Y_f}{Y_s} = e^{\rho s} \left( 1 + \frac{1}{\alpha} \left( \frac{1-p}{p} \right)^\beta \right)^\gamma \]

If potential failure is ignored, by setting \( p = 1 \), \( Y_s \) should provide a compensation over and above the given and guaranteed \( Y_f \) that derives from earnings postponement assessed at the constant income elasticity of utility \( \gamma \). Otherwise, there should be also compensation for risk of failure, depending on parameter values \( a \) and \( \beta \). “Subcertainty”, \( a < 1 \), raises this compensation relative to “certainty” at \( a = 1 \). The second specification only makes sense if the wage for failures is a fixed proportion of the wage for graduates, determined perhaps by their fixed relative skill.

Note that these equations can simply be estimated, and be tested against each other. They require extension of the standard Mincer earnings function for highest level (or years) of schooling accomplished with the probability of failing that level of education, and comparing results for the different specifications. They imply that part of the returns to education should not be ascribed to compensation for postponing earnings, but to compensation for risk.

The risk of drop-out implies a potential difference between ex ante and ex post return to education (assuming marginal returns to education are not constant). If returns are calculated only for those who completed the education, as in the Mincer equation for highest level of education completed, then the estimated ex post return is biased compared to the ex ante return that individuals have to use as a beacon for their decision on enrolment. Ex ante, they should take into account the probability of ending up with the lower earnings of a drop-out (although it is also conceivable that they drop out because they discover halfway school that their alternative is better than they anticipated). If one is interested in the ex ante rate of return, the expected benefit that guides schooling decisions, one should consider the returns for the entire cohort that embarks on the particular education. Ex ante returns for starting an education are given by the entire lifetime earnings distribution of those who started the education, no matter at what exit they will leave the school system. The ex ante rate of return is found from equating lifetime earnings when not entering the education to expected lifetime earnings from all the potential exits once one has entered (i.e. the ex ante probabilities for the exits and the earnings streams associated with each exit).

An example of estimating ex ante returns (and perhaps the only one) is Altonji (1993). Altonji focuses on continuing to college after completing high school. Mirroring the remarks made above, he quotes Weisbrod as saying that the return to the first year of college is not the
difference between 12 and 13 years of education, but the difference between 12 years and the probability distribution of outcomes when continuing into the 13th year. According to his estimates, the ex ante return for men to attending college (in the US in the 1970’s) is 2.8%, while the ex post return to attending college for less than two years is −0.61%. For women, the ex ante return is 8.5% and the ex post return is 7.4%. Ex post returns are found from regressing wages on education level attained, in detailed categories allowing non-linearities in the returns to manifest themselves. Ex ante returns are estimated by estimating probit equations for education level obtained for a given cohort of high school graduates and combining the predicted probabilities with the estimated school dummy coefficients in a wage equation (and their implied lifetime earnings), and then solve for the rate of return.

4.2 modelling the uncertainty

As a consequence of uncertainty in future earnings, the magnitude of the investment in education cannot be determined with certainty. The investment is mostly earnings foregone, but they cannot be predicted with certainty. This makes it also complicated to determine the risk of the investment as the variance in the rate of return. If you buy a security for $1000, the standard deviation in the rate of return is simply the standard deviation of the returns divided by 1000. But if you don’t know how much you will be investing, it is hard to predict the rate of return.

The uncertainty causes no problems if an individual’s relative position in the earnings distribution does not shift over time and between alternatives. Suppose, an individual will always be at the z-th percentile of an earnings distribution for any given education and experience. Then, the rate of return calculated from the means of the distributions is a proper indicator of ex ante returns, except for a correction for differences in variance of the distributions. When the distributions do not simply differ in mean and standard deviation quantile regressions would also be informative. However, the individual’s position may vary with experience, and with education. If the individual moves around in the distribution during her career, mean earnings are no good indicator of ex ante returns. Hence, checking on individuals’ relative position in an earnings distribution over their career, from simply considering transition matrices in percentile positions, in panel data, provides useful information, both on risk in the rate of return and on potential bias in its mean.

In Human capital (1964, third edition), 195-204, Becker applies the general formula for the variance of a product to the rate of return to education. He stylises the rate of return to returns over cost, which makes the variance dependent on four parameters: the variation in (opportunity) costs, the variation in returns, the correlation between returns and costs and the correlation between returns in different periods. Based on cross-section evidence then available, Becker concludes that the coefficient of variation in the rate of return will be at least one and probably higher. Presumably because these calculations are based on cross-section data, Becker concludes: “So great is it that an individual can be only loosely guided by the gain of his cohort, and has to place considerable weight on his own situation, and hope for the best”. With the now more abundantly available statistical information, more reliable estimates are no doubt feasible. A more precise calculation of ex ante variance would follow from the approach indicated above: compare lifetime earnings from not entering an education to the probability distribution of lifetime earnings when one does enter the education, with probabilistic exits and stochastic earnings waiting after each exit. In the latter case, there is no simple explicit solution.
Note that the variance in the ex ante rate of return cannot be read from the standard deviation of the estimated regression slope on schooling in a Mincer equation. This standard deviation reflects the reliability of assigning differences in earnings to differences in schooling when many other variables, as well as measurement errors in the dependent variables, play a role. It would indicate an individual’s ex ante uncertainty when assessing the returns from schooling (using regression analysis), but it ignores variation in the cost and the correlations between cost and returns. A similar remark holds when the rate of return to schooling is estimated in a random coefficient model (Harmon, Hogan and Walker, 2001): it reflects variance in the earnings differences attributed to schooling differences, but ignores the other parameters determining the rate of return.

With uncertainty about even the amount invested, one may doubt whether using a model of individual choice based on mean and variance of the rate of return to an investment is a proper approach. The rate of return follows from comparing two earnings flows, both uncertain. Hence a choice model simply considering two alternative flows may be more adequate than a model in which these two flows are compressed in measures of their difference.²

5. INVESTMENT IN HUMAN CAPITAL UNDER RISK

5.1 The Levhari-Weiss model

The model developed by Levhari and Weiss (1974) is the standard model for human capital decisions under uncertainty. Eaton and Rosen (1980) augmented it for analysis of taxation and Kodde (1985) used it as the basis for empirical work on The Netherlands. The model has two periods, a schooling and a working period, time is allocated in the first period between schooling and working, leisure has no utility. There is uncertainty in the returns to schooling.

The basic model is constructed as follows. Utility only depends on consumption $C_1$ in period 1, $C_2$ in period 2; assuming expected utility maximisation, the objective function is $E\{U(C_1, C_2)\}$. In period 1, time is allocated between hours worked $H_W$, and hours spent investing in human capital (schooling), $H_S$, both expressed as a proportion of time endowment; the constraint implies that we may write $1-H_S$ for $H_W$. The first period wage rate is given as $W_1$, initial assets are given as $A$. Second period earnings are a function of the time invested in period 1,

\[ W_2 = h(H_S, x) \]

where $x$ is the random variable representing the state of the world that is revealed at the beginning of period 2. We assume $h_x = \frac{dh}{dx}$ positive: a higher value of $x$ refers to a better state of the world. Hence, uncertainty is brought in as uncertain future pay-off to this period’s investment of time, evaluated at a known, non-stochastic wage rate. We can associate hours spent investing (schooling) as the share of a fixed length first period going to school, e.g. the

² Note, for example, that the well known Willis and Rosen model of self-selection into post high school education is not cast in terms of a rate of return to the investment.
share of the first ten year interval from age 15 to age 25. It is a slightly odd specification, as it implies that if you go to school for five years (up to age 20), you will only get your returns at age 25, when the second period commences. Nevertheless, in some sense the uncertainty covers both input risk and output risk in Levhari and Weiss’ sense (not knowing the quality of your inputs and not knowing what your inputs will yield). Levhari and Weiss hint at x as a measure of individual ability, not known when investment decisions have to be made, revealed when entering the second period.

With a given rate of interest r, we can express the intertemporal budget constraint as an expression for $C_2$.

\[ C_2 = h(H_s, x) + (1 + r)(A + w_1(1 - H_s) - C_1) \]

Maximizing $E[U(C, C_2)]$ with the instruments $H_2$ and $C_1$ requires as first-order conditions

\[ E \left[ \frac{dU}{dC_1} - (1 + r) \frac{dU}{dC_2} \right] = 0 \]

\[ E \left[ \frac{dU}{dC_2} \left( \frac{dh}{dH_s} - (1 + r)w_1 \right) \right] = 0 \]

An immediate effect of uncertainty is the absence of separability of consumption and time allocation that holds under certainty, with a perfect capital market. If there were no uncertainty, condition (16) requires the inter-temporal marginal rate of substitution of consumption to be equal to the rate of interest. With non-satiation, the second condition would state, under certainty, that investment in human capital should proceed until marginal benefit equals marginal cost. The two conditions are independent from each other. But with uncertainty, we cannot eliminate the marginal utility of consumption from condition (17). Investment in human capital is now inevitably related with consumption smoothing, and conversely, of course. The term in square brackets in (17) is the difference in second period returns to an hour's time in period 1 spent investing in human capital or spent working with the proceeds saved: it's the difference in return to an hour's effort invested in human or in non-human capital.

While under certainty investment behaviour generates equality between marginal returns to human and non-human capital, this no longer holds as a general rule under uncertainty. Rewrite (17), with simplified symbols $U_2$ and $h_s$, as

\[ \frac{EU_2 h_s}{w_1 EU_2} = 1 + r \]

and subtract expected return on human capital from both sides

\[ \frac{EU_2 h_s}{w_1 EU_2} - \frac{Eh_s}{w_1} = (1 + r) - \frac{Eh_s}{w_1} \]
Then, the sign of the difference between marginal return on non-human and human capital depends on the sign of the left-hand side, which can be rewritten as

\[
(20) \quad \frac{E(U_2)E(h_s) + \text{cov}(U_2, h_s) - Eh_s}{w_1EU_2} = \frac{\text{cov}(U_2, h_s)}{w_1EU_2}
\]

Thus, the expected marginal return to investment in human capital is greater than the return to non-human capital investment if \( \text{cov}(U_2, h_s) < 0 \). Kodde (1985, Chapter 7) shows, using a second-order Taylor expansion\(^3\) around \( E(x) \), that this covariance term can be written as

\[
(21) \quad \text{cov}(U_2, h_s) = U_{22} h_s h_{ss} s^2 - [U_{222} h_s^2 h_{ss} + U_{22} h_s h_{ss} s^2] / 4
\]

where \( s^2 \) is the variance of \( x \). If \( h_{ss} = 0 \), as Levhari and Weiss implicitly appear to have assumed, the sign of \( h_s \) determines the sign of the covariance. \( h_s \) is the sensitivity of the marginal return of schooling to the stochastic shock in the wage rate. With an additive specification of the returns, \( w_2 = g(H_s) + x \), the derivative is zero, the covariance is zero, expected return from human capital equals return to non-human capital (in equilibrium) and risk has no effect on human capital investment. If the marginal return to schooling increases in good states (\( h_s > 0 \)), the covariance is negative (from \( U_{22} < 0 \)), expected return from schooling surpasses the rate of interest and investment responds negatively to (increasing) risk \( s^2 \) (from decreasing marginal returns to schooling).\(^4\) Conversely, if the marginal return to schooling decreases in good states (\( h_s < 0 \)), investment in schooling is increasing in risk \( s^2 \).

With a multiplicative specification for risk, \( w_2 = xg(H_s) \), \( h_s > 0 \) and investment reacts negatively to increasing risk.

The Levhari and Weiss model then, has a clear conclusion. Increasing risk reduces investment in human capital if good states of the world generate higher marginal returns to schooling. If \( x \) is ability, the condition entails higher marginal returns for higher ability individuals. This will show up as a positive correlation between marginal and average returns to schooling (higher ability individuals have higher average returns). Levhari and Weiss point out that increasing risk can be tested: the required positive correlation between marginal and average return to human capital holds if the variance of earnings increases with the level of investment in human capital (o.c., p. 954). In fact, this appears to be an empirical regularity.\(^5\) Note that these results refer to the expected marginal returns to human capital. Levhari and Weiss (1974, p. 954) show that risk aversion (a concave utility function) is sufficient for the average return to human capital to surpass the safe return to non-human capital: investment behaviour ensures a positive average risk premium on the risky investment.

Still assuming \( h_{ss} = 0 \), a little rewriting shows

\(^3\)Applying a third-order Taylor expansion would bring the model in line with the empirical tests for risk compensation in wages in Hartog and Vijverberg (2001); cf section 6 below.

\(^4\)Intuitively, this is remarkable. With ability \( x \) unknown, investment decreases if one knows that if ability happens to turn out high, the benefit from schooling is higher. Empirically, one observes that individuals with higher measured ability (IQ) have higher schooling levels.

\(^5\)See for example Low and Ormiston (19..) cited above.
\[ (22) \quad (1 + r) - \frac{Eh_x}{w_1} = \frac{\text{cov}(U_2, h_x)}{w_1 EU_2} = \frac{U_{22} h_x h_{2x} s^2}{w_1 EU_2} = \frac{w_1 U_{22} h_x h_{2x} s^2}{w_2^2} \]

\( w_1 U_{22}/EU_2 \) is a measure of relative risk aversion, evaluated at the first period wage and the expected value of \( U_2 \). The larger the degree of risk aversion, the greater the gap between returns to non-human capital and expected returns to human capital, and the greater the effect of increasing risk. Note that the return to non-human capital has been assumed non-stochastic. Thus, in a stationary world, where individuals behave in the Levhari-Weiss mode, and where post-school wages directly reflect productivity, the wage premium for given risk (the square of the coefficient of variation) is increasing in risk aversion. If one estimates a wage equation expanded with the measure of risk, the estimated coefficient would be the coefficient of relative risk aversion as defined in (22); the measure of risk could be taken from post-school earnings, as the formula uses the transformation of risk in \( x \) to risk in second period wages (however, relative to first period wages!).

As a point of reference, we note that the pure human capital model under conditions of certainty, and with a perfect capital market, predicts the following comparative static results. An increase in initial assets (or family wealth) has no effect on investment or returns, because of the separability of consumption and investment. An increase in the market rate of interest, in the direct cost of education (tuition, books, etc), in the opportunity cost (the unskilled wages) and a decrease in scholarships and grants \textit{a fonds perdu}, all reduce investment in human capital and increase the marginal rate of return to human capital (assuming declining marginal returns to investment). An increase in the marginal productivity schedule increases investment and returns. These conclusions hold for homogenous labour supply groups, of given ability, and for homogeneous schooling, that is, in the absence of comparative advantage effects.

The Levhari-Weiss model, for investment under uncertain returns, generates predictions under the additional assumptions of increasing risk (returns up in better states of the world) and decreasing absolute risk aversion. Increasing initial assets (family wealth) now increases investment in human capital (because higher wealth reduces risk aversion). An increase in the market rate of interest leads unambiguously to a lower investment for a net borrower (as the substitution effect is re-inforced by a negative income effect), while the prediction for a net saver is ambiguous (as the income effect now counters the substitution effect). Kodde (1985, p. 136) derives that under an intertemporal additive separable utility function and decreasing absolute risk aversion, an increase in the direct cost of education or a decrease in schooling grants reduce the level of investment (the negative substitution effect is reinforced by moving away from the risky investment if real income is reduced).

Eaton and Rosen (1980) analyse the effect of taxation in the Levhari-Weiss model. They note first of all that under a constant marginal tax rate on labour earnings the tax has no effect on human capital investment (as forgone and future wage are affected by the same proportionate reduction). This, of course, implies that the tax is fully borne by labour, but this conclusion depends crucially on fixed labour supply in the post-school period. A proportional tax on interest (i.e. a constant marginal tax rate) increases human capital investment. A differential tax treatment of human and non-human capital income leads to a shift away from non-human capital. Jointly, these results imply that a proportional \textit{income} tax stimulates human capital investment, as it implies that interest income is taxed, while the gap between foregone labour income and future labour income is not. Under uncertainty, the results are not so clear-cut,
and generally ambiguous. Only with constant absolute risk aversion is the effect of proportional labour taxation on human capital investment unambiguously positive. Endogenising second period labour supply further complicates the analysis, without yielding unambiguous results.

As pointed out by Jacobs (1998), the Levhari-Weiss model can be solved explicitly for some specific assumptions. In particular, with the utility function

\[ E[U(C_s)] = \frac{1}{\beta} e^{-\gamma C_s} \]

where \( \gamma \) the constant degree of absolute risk aversion, \( \beta \) is the discount factor, and \( H_1 \) is hours of leisure in the first period (leisure in the second period is ignored), with multiplicative risk \( w_s = x h(H_s) \), with \( x : N(1,s) \) human capital has a normal distribution, with expectation \( h(H_s) \) and variance \( h(H_s)^2 s^2 \). With these specifications, the equivalence between EU and MS models holds (see Appendix). The first order condition for a maximum in \( H_s \) is then

\[ \frac{dh}{dH_s} = (1 + r)(w_i + b) + s^2 h(H_s) \frac{dh}{dH_s} \]

This is the familiar equality of marginal returns to marginal cost, \( b \) is direct outlays on books and tuition, the latter now augmented by the cost of risk. Familiar predictions also easily return. The equilibrium condition has been graphed in Figure 1.

The second-order condition requires marginal cost to be above marginal revenue at investment beyond the optimum level, and this implies that under uncertainty the optimum investment, \( H_s' \), must be lower than under certainty, \( H_s' \) (when the risk premium on marginal cost is absent). Clearly, the optimal investment level increases for lower interest rate, lower opportunity wage, lower direct cost, higher grants, lower degree of risk aversion. The impact of uncertainty, as compared to the optimum under certainty, is smaller if risk aversion is lower.

Kodde (1985) has empirically tested the predictions of the Levhari-Weiss model. His dataset on enrolment in post-secondary education includes subjects’ estimates of foregone and future earnings. Earnings risk is measured as the difference between highest and lowest future monthly earnings (the difference between log earnings has a mean of 0.65 and a standard deviation of 0.32). It turns out that enrolment is significantly positively related to the earnings difference, thus violating the prediction that increasing risk decreases investment in human capital. As usual, there is a caveat. The testing is not on individuals confronted with an increase in earnings risk, but across individuals with different expectations of their individual risk. Hence, the coefficient may reflect other unobserved personal characteristics than just a difference in the risk they are confronted with.

---

\( ^{6} \) Note that compared to certainty, the marginal cost is increased by the fraction \( \rho \sigma^2 h \) of marginal productivity \( \partial h / \partial H_s \). This fraction is increasing in \( H_s \) as \( \partial (\rho \sigma^2 h) / \partial H_s = \rho \sigma^2 \partial h / \partial H_s > 0 \).
The Levhari-Weiss model is subject to severe restrictions as a description of the choices facing the individual. The student is supposed to be confronted with the option of spending any amount of hours on producing human capital, with uncertain effect however. The uncertainty may reflect, as Levhari and Weiss call it, input uncertainty and output uncertainty. The former reflects the effect of unknown ability: the student may be more or less successful in her schooling career. The latter reflects the vagaries of the labour market, where future value of investment is hard to predict. As a specification of output uncertainty, the model is attractive. As a specification of input uncertainty, it seems too far removed from the institutional features of educational systems. Moreover, it does not seem to do full justice to tastes of students, neither for leisure, nor for the intellectual adventures of an education. And recognition of capital market imperfections also seems to be called for (as the authors admit in their final section).

5.2 Williams’ model

Remarkably, the model developed by Williams (1979) is almost neglected as a model for human capital investment under uncertainty, while in fact it is much more general than the routinely cited paper by Levhari and Weiss. The model by Williams integrates investment in human and in non-human capital and thus precedes a literature that is only recently developing.

Adopting uniform symbols throughout this paper as much as possible, Williams’ model considers marketable skills (human capital) $K$, defined as maximal current labour income, and time allocation, in percentages, $H_s$ to education, $H_l$ to leisure and hence $(1-H_s-H_l)$ to work. Time is continuous, from 0 to $T$, when death strikes with certainty. Work in any given period is paid at rate $w_t$, with present wage known and future wages uncertain. Individuals can invest their wealth $V$ in a single riskless asset at return $r$, in $N$ risky marketable assets or spend it on consumption $C_t$. Human capital depreciates at an unknown, stochastic rate $d$. Human capital is produced linearly as $H_s K_t$.

At the beginning of period $t$ the individual knows his human capital $K$, his non-human wealth $V$, his wage rate $w_t$. Depreciation rate $d$, education productivity $\beta$, future wages $w$ are all characterised by a Wiener process. Depreciation $d$ has a lognormal distribution, with mean $\mu_d$ and variance $\sigma_d^2$, education productivity has a lognormal distribution with mean $\mu_\beta$ and variance $\sigma_\beta^2$, future wages relative to current wage are distributed lognormally with mean $\mu_w$ and variance $\sigma_w^2$. Depreciation $d$ and school productivity $\beta$ have covariance $\sigma_{d\beta}$. The $N$ risky marketable assets exhibit prices lognormally distributed over time, with mean vector $\mu$ and variance-covariance matrix $\Sigma$. The marketable assets have covariance matrix over time with contemporaneous adjustment in wages $\Sigma_w$. All other covariances are assumed zero. Depreciation and school productivity are unrelated to wage adjustment and to asset prices.

Williams derives optimum conditions for consumption $C$, leisure $H_l$, schooling $H_s$ and investment in marketable assets, in shares $X$ of the total risky portfolio, from maximising discounted lifetime utility, where utility depends on consumption, leisure, human capital stock and bequest. The optimum conditions are expressed on the indirect utility function $J$, maximal discounted lifetime utility as a function of $K$, $V$ and time $t$. 

16
For consumption and leisure, the optimum conditions are, at any time $t$,

(25) $U_c/J_v = 1$

(26) $U_c/U_l = 1/K$

The first condition holds the rate of substitution between consumption and financial wealth (savings) constant, the second condition holds the rate of substitution between leisure and consumption equal to the inverse of human capital. For investment in human capital the condition is:

(27) $H_s = \left(-\frac{J_k}{J_{kh}V}\right)\left(\frac{\mu - (J_s/J_k)}{s^2}\right) + \frac{s_{\sigma}}{s^2}$

and for investment in risky assets

(28) $VX = \left(-\frac{J_s}{J_{sv}}\right)\Sigma^{-1}(\mu - rI) - \left(\frac{J_k}{J_{sv}}\right)\Sigma^{-1}\Sigma_w$

The structure of these equations is quite clear. Note, first of all, that investment in human capital does not respond to any financial parameter, other than through effects on the derivative of the indirect utility function. The first term in (27) is the inverse of relative risk aversion for investment in human capital, evaluated on the indirect utility function, and will be positive. Human capital investment reacts positively on the expected rate of return corrected for the rate of substitution between human and non-human wealth. More non-human wealth will reduce its marginal indirect utility and thus increase investment in human capital. Investment in human capital reacts positively to the covariance between depreciation and school productivity. If high depreciation comes along with high school productivity, individuals invest more. With negative covariance they invest less, as the two components act like each other’s insurance. The effect of increased risk, as a greater variance in school productivity, is to reduce investment in human capital, unless the covariance between depreciation and productivity is highly negative and risk aversion is very strong.

Equation (28) specifies the optimal portfolio for risky assets. The first term is the standard investment component, with absolute risk aversion determining the response to the return above the safe return on the riskless asset, variance weighted. The second part is the asset portfolio response to the labour market. If the covariance between marketable assets and wage adjustment is zero, the portfolio does not respond to the wage risk. If human and non-human wealth are substitutes $(J_{sv}<0)$, investment in an asset responds negatively to covariance between wage and asset: there is less investment in an asset that covaries positively with wage adjustment, and more in an asset that covaries negatively with wage adjustment.

Williams also draws out conclusions for observable lifecycle patterns of consumption, leisure, investment etc, and derives an explicit solution for a specific example of preferences. Most interestingly, he develops the case where time allocated to education also improves information on ability. Individuals start out not knowing the mean of their educational productivity $\mu$ nor the variance $s^2$, only their distribution and update their beliefs from
observing labour income and hence human capital at the end of each time interval. In the
optimal investment in human capital, the individual now uses his posterior mean for
educational productivity, which implies that lucky draws spell optimism about ability and
lead to additional investment. However, the added variance from not knowing his true ability
reduces his investment. Moreover, investment is increased to generate information about his
ability. Unambiguous predictions cannot be made, as they depend on the balance and
evaluation of countering effects.

5.3 Conclusions on risk and human capital investment

The seminal paper by Levhari and Weiss predicts lower investment in human capital when
the risk in the returns increases. Williams’ model predicts that human capital investment does
not respond to increased risk in future wages, as this will be absorbed in portfolio adjustment.
However, it does respond to increased risk in school productivity and to the covariance
between depreciation and productivity of schooling.

Clearly, testing the sensitivity of human capital investment to risk is an interesting target for
empirical research. One might use cross-section international or inter-regional data, using
wage dispersion for given education as the measure of risk and assess how demand for
education responds. The foreseeable risk in depreciation may be hard to measure. The risk in
the productivity of schooling may be measured from variability in school performance
measures, although they may be affected substantially by self-selection. As far as I know,
Kodde (1985) is the only empirical test of the impact of risk on human capital investment.

6. ON THE PREMIUM FOR EARNINGS RISK

6.1 Introduction

In his famous discussion of the structure of wages, Adam Smith lists five principal circum-
stances which give rise to wage differences between employments:

* the agreeableness of the job,
* the cost of training,
* the variability of employment,
* the required degree of trust in the employee,
* and the probability of success.  

The first is generally acknowledged in empirical work, often with reference to the theory of
hedonic prices. The second has given rise to an enormous literature and a truly canonical
specification of an earnings function: the log of wages related to years of education,
(potential) experience and its square. The third has been acknowledged in a modest body of
empirical work, the fourth does not seem to have survived to the present day as something of
importance. The fifth is probably widely acknowledged as relevant, but there is only a modest
literature.

---

7 Smith (1776/1976, Chapter X, Part 1, pp 202-213)
The purpose of this section is to take a closer look at Adam Smith's Fifth Commandment: the relation between wages and risk. In the examples that Smith discusses (lawyers and entertainers), there is interaction with other principles (in particular the first), but still some interesting specific hypotheses shine through. Smith starts out by noting that the probability or improbability of success differs very much between occupations. He then notes that occupations in which success is uncertain are underrecompensed. If compensation were to work as a fair lottery, "those who draw the prizes ought to gain all that is lost by those who draw the blanks". While such a fair lottery would not at all compensate for risk, Smith argues that the rewards are even less. He explains this from compensation in other forms: public admiration and the common overestimation of the chance of gain. The latter effect is strongest in youth: "The contempt of risk and the presumptuous hope of success are in no period of life more active than at the age at which young people choose their occupations." It has been argued that this holds in particular for artistic jobs in which many young entrants try their luck, depressing wages while a selection process starts to expel the less successful, leaving high rewards for those who do succeed. However, the argument is not supported by the evidence (Filer, 1986).

An effect of risk on wages has been established in several studies. The common procedure is to measure risk from the variance of the residuals in an earnings function. In cross-sections, this is the standard Mincer specification (with education, experience, etc), and the residuals are collected by occupation or industry. In longitudinal data, the risk may be estimated from individual residual variation over time. McGoldrick and Robst (1996) discuss several of these studies, with the common finding that risk is indeed compensated in individual wages.

McGoldrick (1995) specifies a simple occupational choice framework with a worker utility function in wage, wage risk, wage skewness and non-wage job characteristics. Workers dislike risk R (earnings variance), but like skewness K (kapital): they appreciate small probabilities of receiving higher earnings. The wage function is specified in the same variables. The first-order conditions for occupational choice predict a positive wage coefficient for risk (it is equal to the marginal rate of substitution between risk and wage, \(- \frac{U_K}{U_W}\)) and a negative coefficient for skewness (it is equal to the marginal rate of substitution between skewness and wage, \(- \frac{U_K}{U_W}\)). The utility function is directly specified in the variables W, R and K, and hence, is an extension of the MV approach. King (1974) found mean income in an occupation to be positively related to the standard deviation and negatively to the skewness, also when adding mean ability levels in the occupation.

McGoldrick and Robst (1996) extend this basic framework with worker mobility. They simply add the degree of mobility to the utility function, and confess ignorance on the sign of the derivative. Wages are assumed to depend on the degree of mobility, including interaction between mobility and risk compensation. They solve the optimal job choice problem as maximizing utility subject to the wage function, and derive first order conditions on wage, risk, mobility and non-wage job characteristics. They then derive that the total compensation for risk (including the interaction effect) should be positive and that the total compensation to mobility (including the interaction) should also be positive, assuming the marginal utility of mobility is negative. As predicted, the compensation for risk is lower if mobility is higher.

---

8 "It makes a considerable part of that reward in the profession of physic; a still greater perhaps in that of law; in poetry and philosophy it makes almost the whole."
(both risk and mobility are positively compensated, and the interaction term is negative). The probability of mobility is higher if risk is higher (the authors estimate a two-stage model for mobility and wages). The model is not quite satisfactory due to the treatment of mobility, in particular in the utility function.

Some studies have also brought in individual's risk attitude. Bellante and Link (1981) use an index of risk aversion developed in the Michigan PSID data, a combination of data on insurance behaviour, use of seat belts and drinking and smoking behaviour. They find that more risk averse individuals are more likely to have a job in the public sector. Feinberg (1981) uses the same variable in an earnings function with compensation for risk (income variance) and finds that more risk averse individuals get a higher compensation for risk.

Olsen, White and Shefrin (1979) specify a model that mimics the situation for American students considering college training after finishing high school, including the scholarship system allowing to take up loans at specific conditions. They find that accounting for risk actually increases the return to college education, but the effect is very modest.

An elegant expression to correct estimated rates of return for risk has been developed by Weiss (1972). Assuming a lognormal wage distribution and a utility function with constant relative risk aversion, we can derive

\[ \ln Y = 0.5 s^2 + ds/(1 - \gamma) \]

Where \( d \) is the discount rate and \( \gamma \) the coefficient of relative risk aversion. A rate of return in private industry for a Ph.D. program of 7.8% is reduced to 5.7% with risk aversion of 0.2 and to 0.7% with risk aversion of 0.6. Hause (1974) has pointed out that Weiss' correction is mostly sensitive to postponement of obtaining an income, rather than risk aversion. This has to do with confounding risk aversion and time preference (cf section 2.1 above). Deviations of an income stream over time from a constant income imply utility loss even under certainty, due to diminishing marginal utility. Weiss' formula only holds if there is no access to a capital market for smoothing income flows. Under perfect capital markets, the proper maximand for dealing with risk is the expected value of the discounted income stream, not the expected value of discounted utilities: under perfect capital markets, individuals have the total expected income stream at their disposal.

Low and Ormiston (19..) use the MV approach. With utility function \( v(\mu, \sigma) \) in mean and standard deviation, schooling choice maximizes lifetime earnings

\[ V(s, \rho) = \int_0^T e^{-\rho(s + t)} v(\mu, \sigma) dt \]

The internal rate of return \( \rho \) is defined by \( \partial V(s, \rho) / \partial s = 0 \). Specifying the utility function as

\[ v = \mu^{1-k} (\sigma + 1)^{-k} \]

\( \rho \) can be solved to
where $\mu_s = \partial \mu / \partial s$ and $\sigma_s = \partial \sigma / \partial s$. With $k=0$, we have risk-neutrality and $\rho$ equals the relative marginal returns on schooling, with risk aversion ($k>0$) there is a correction for the disutility of risk. Hence, just as in Weiss' approach, if we knew the degree of risk aversion, we could calculate the risk adjusted rate of return. This is indeed what Low and Ormiston do. They estimate an earnings function with parameterised heteroscedasticity.

$$Y_i = X_i \alpha + X_i \beta e_i$$

Results from the NLS data for adults aged 27-39 in 1981 show dispersion to increase in education for both men and women, to be decreasing in tenure for men and increasing for women and to be increasing in IQ for women, and not significant in IQ for men. Risk adjustment on the estimated rate of return may be substantial, but of course is entirely dependent on the assumed degree of risk aversion. Moreover, the specification of the utility function has the drawback of implying a risk premium that is increasing in $\mu$, i.e. increasing absolute risk aversion. Empirically, this is highly implausible.

6.2 Hartog and Vijverberg

In Hartog and Vijverberg (2001) we have developed and estimated models for risk compensation in wages and subjected them to extensive empirical testing. We find support for a positive effect of risk and a negative effect of skewness on wages.

The Hartog and Vijverberg model was developed employing EU. It is straightforward however, to apply MV. Let the individual face an income distribution with mean $Y_o$, standard deviation $s_0$ and skewness $\gamma_o$, when going to work and an income distribution ($Y_s$, $s_s$, $\gamma_s$), after going to school for $s$ years. If risk involves a single draw for the entire lifecycle income profile, drawn right after $s$ years of schooling, and if $Y_s$ is assumed to adjust to equality of lifetime utilities, we would require

$$\int_0^\infty U\{Y_o, s_o, \gamma_o\} e^{-\rho t} dt = \int_s^\infty U\{Y_s, s_s, \gamma_s\} e^{-\rho t} dt$$

Which can be solved with a first-order Taylor expansion for the schooling mark-off as

$$M_s = \left\{1-e^{-\rho s}\right\} \frac{1}{E_u} + \frac{s_s-s_o}{Y_s} \mu \gamma + \frac{\kappa_s-\kappa_o}{Y_s} \mu \gamma$$

---

9 The risk premium that compensates for increasing risk by increasing $\mu$ is found as the slope of the indifference curve for $v$. Defining this slope as $\Pi_a$ (for absolute risk premium, see below) we have $\Pi_a = -d\mu / d\sigma = -k \frac{\mu}{1-k} \frac{1}{\sigma + 1}$. Thus, the constant utility risk premium is increasing in $\mu$ and decreasing in $\sigma$, a very unattractive specification.
where $\varepsilon_y$ is the elasticity of utility to income, evaluated at $Y_s$, with $\sigma_y = \kappa_y = 0$, where

$\mu_{ys}$ is the marginal rate of substitution of mean income for dispersion, $-u_1^y/u_1^s$ and $\mu_{ys}$ is the mrs of mean income for skewness, $-u_1^y/u_1^s$. This is almost identical to the reduced form model estimated in Hartog and Vijverberg, such as equation (26) in section 2.3. Note that we assume that $Y_s$ adjusts to equilibrium which $s_s$ and $\kappa$ are kept constant. They may be so because of technological constraints, or perhaps to generate proper incentive structures. One might also conceive of mechanisms where the firm distribution adjusts so as to generate the equilibrium variance structure in the sector.

We can also apply CPT, if we use the continuous version.

### 6.3 Labour markets and portfolio choice

Williams (1979) is an integrated model of human capital investment and portfolio choice for financial investment. Recently, this literature has been further developed from the perspective of portfolio choice models, where the assumption of a given financial wealth has been replaced by the assumption of risky labour income. An example is Davis and Willen (2000). To them, the significance of risky labour income is the breakdown of the classical two-fund separation theorem. Because the difference in cost of risk between investors is independent of the asset under consideration, all individuals optimally hold the same relative portfolio, that is, the same percentage contribution of different assets. In fact then, a single mutual fund, and a riskless bond, will allow all investors to obtain their optimal portfolio. This no longer holds if individuals’ starting positions are different: they start out with a particular investment in human capital and the risk that comes with it. If labour incomes correlate with asset returns, different portfolios mean different things to different investors, because the correlation with their labour income positions them differently. However, we can invert their approach and consider whether labour market risk can be hedged in the stock market. This is only feasible if the covariance between labour income and asset returns is non-zero.

Figure 2, from Davis and Willen (their fig 1, replacing X by A), depicts the standard analysis. For a risky asset with some expected return, the expected benefit (in Euro) from investing an amount $A$ is given by the straight line Benefit. The dotted cost curve indicates the investor’s appreciation of risk. For the given variance of the returns to the asset, the investor wants a return in terms of a mark-up over the return on a riskless asset that increases in the amount invested. It’s the demand price for a given risk, as a function of the amount invested. Clearly, the optimum investment is where the distance between the curves is largest, or where their slopes are equal.

Now assume, the return to the asset is negatively correlated to labour income. In Davis and Willen’s example, the individual may work at Chrysler Automobile Works and consider investing in shares of competitor Honda. When Honda does well, labour earnings at Chrysler will be under pressure. Now this shifts the cost curve to the right and down: buying Honda stocks reduces the variability of total income, as the bad draw of labour income at Chrysler will be compensated by the good draw of Honda stock. The investor is willing to pay a premium on Honda stocks, i.e to accept expected returns below a safe return, as his risk is being reduced. Of course, this only works up to a point, so after a while the cost curve turns up again. Davis and Willen call the distance from the origin to the minimum of the cost curve
OE endowed risk. It measures the shift of the origin, the amount to be invested to reach
minimum cost of risk. The optimum position would again be found at equal slopes of the two
curves. The distance between the minimum cost investment (at E) and the optimal investment
(at A) is called the desired exposure. The investor buys OA and is effectively exposed to EA,
as OE is already covered by labour earnings risk. (D&W, fig 3)

Now consider the case of positive correlation between labour earnings and the asset, i.e. the
Honda share for a Honda employee. In this case, the cost curve shifts up and to the left
(D&W, fig 4, after correction). Any shock to the Honda share is aggravated by a shock to
labour earnings at Honda. The variability is aggravated and the required premium goes up.
By going short in Honda shares, however, the Honda employee can compensate the effect of
a bad draw on labour earnings by benefiting from the low price of Honda stock. It is
important to note that individuals are supposed to have unlimited access to the capital market.
Hence, if they want to go short, or want to buy shares, the bank is always willing to support
them. This, of course, is a very strong assumption. Workers can always exploit any non-zero
covariance between wages and stock prices. Hence, whatever the endowed risk, they can
always reach their optimal position, because they are not refrained from exploit the non-zero
covariances. Thus, reversing the conclusion: if the covariance between wages and the stock
market is zero, they cannot hedge in the stock market against labour market risk. Neither can
they if they lack the wealth and the bank support to finance any desired position in the stock
market. Davis and Willen’s analysis assumes away the latter constraint, which appears utterly
unrealistic. Their empirical research probes the covariances, and these are interesting enough
for a further look.

To assess empirical relevance, and options for hedging away labour market risk, Davis and
Willen consider three variables, correlation, variability and persistence. They cite three
studies: Two of these find that the correlation between labour income shocks and equity
returns rises with education. A third finds a positive correlation between occupation-specific
shocks from occupation-specific year dummies. The standard deviations of these
shocks are between 2 and 7% of income. Persistence is measured with a second-order ARMA
model on the first difference of the occupation-year effects. With the estimated ARMA
parameters, they calculate the present value multiplier, as the effect on discounted lifetime
income of a 1 dollar shock to income, at discount rate of 2.5 %. At age 30, for the ten
occupations, they vary between 10 and 27, with outliers of 7 and 40.

The essential parameter is of course the correlation. Now, here the results are highly negative
and quite disappointing for advocates of hedging labour market risk in the stock market. To
quote: “We find little evidence that occupation-level income innovations and aggregate
equity returns are linearly related in annual data over the period 1968 to 1994.” (o.c., p 15)
For none of the ten occupations is there a statistically significant relationship. By using
industry portfolios rather than a value-weighted market portfolio, and matching occupations
to particular industry portfolios (e.g. a portfolio in construction for electricians and
plumbers), they get six correlations around −0.3, and one of +0.22, and the other values in
between. However, if they then use all their information to compute optimal portfolios for
each of these occupations, they find completely unrealistic values. For example, a 40 year old
truck driver should hold a portfolio of 550 000 dollars of 1982, including a short position in
one of three portfolios of 141 000 dollars. It’s easy to agree with the authors’ conclusion that
the resolution of the puzzle (which is mostly a puzzle for investment theory, not for the
analysis of risk in the labour market) must lie in then opportunity cost of investor funds. “If,
instead, investors must borrow at an interest rate that approximates the expected return on
risky assets, the optimal risky asset position is approximately zero when asset returns and
labor income are uncorrelated.” (o.c., p 21). Fama and Schwertz (1977) also find that capital
market returns and labour market returns are poorly correlated. They test a model developed
by Mayers as an extension of the standard Sharp-Lintner-Black model of the capital market
equilibrium, as a relation between expected return on a marketable asset and its risk. As non-
marketable asset they take human capital, and as its return aggregate income from labour and
proprietorship income, as the total return on human capital. When adding the predicted
relations with the non-marketable asset, they find no discernible improvement in the fit of the
capital market equilibrium relationship.

Hence, the conclusion is devastatingly simple: workers must stand on their own feet when
confronted with financial risk in the labour market. They cannot hedge their way out through
the stock market.

6.4 Conclusion on insuring labour market risk

Labour market risk cannot be insured in the way cars and houses can be insured. Moral
hazard and adverse selection have prevented the emergence of firms specialising in labour
market insurance policies. This compels the individual to attempt consumption smoothing by
precautionary saving. A recent literature considers whether the stock market can provide
insurance. The conclusions so far are mostly negative. The models tend to generate
predictions that are patently at variance with the evidence. As Haliassos and Michaelides
(2000) note, earnings of college-educated males have a small positive correlation with stock
returns, while high school dropouts have a marked negative correlation (about –0.25). This
would imply that high school dropouts should invest more heavily in the stock market than
college educated. However, empirical evidence shows a strong positive correlation between
education and stock holding. In an infinite horizon model developed by Heaton and Lucas,
with uninsurable labour income shocks, zero correlation between labour income and stock
returns predicts complete portfolio specialisation in stocks, a quite robust result under
alternative assumptions, but clearly at variance with the evidence. Haliassos and Michaelides
propose fixed entry cost to deter such full specialisation. The important observation is that
most households do not hold stocks in any significant amount, in spite of the recent
emergence of something of an “equity culture” among households in the US and Europe.
Whatever the models may suggest (and so far that literature seems to have more puzzles than
answers), hedging labour market risk to the extent that individuals are indifferent to exposure
to different degrees of labour market risk is not an attractive assumption. Borrowing
constraints, lack of correlation, absence of empirical evidence that individuals do indeed seek
insurance in the stock market, there are strong arguments for assuming that labour market
risk should be compensated in the labour market itself. The link between labour market risk
and portfolio as an interesting issue is mostly an issue of extending models of portfolio
choice, not the other way round, with the stock market a place that can absorb all the
worker’s earnings risk.

The conclusion is simple. Suppose an individual can choose between a job with lifetime fixed
income of 1 million, and a job with equal probability of either 0.5 or 1.5 million. Can he buy
insurance in the stock market so that he will be indifferent between the two jobs/incomes? If so, employers are not compelled to pay a premium to attract supply to the risky job. Suppose he can invest 1 million in a share or fund that with probability 0.5 brings him a profit of 0.5 and with equal probability a loss of 0.5 million, and that the stock market and labour market probabilities are perfectly negatively correlated. So, he may take the safe job, with lifetime income of 1 million, or the risky job combined with the risky stock, also earning him 1 million for sure. But in the latter case, he either must have the wealth to do this (and foregoing a safe return from a savings account, so in fact the return in the stock market must be higher than in the labour market, to include forgone safe returns), or he must borrow it, and also pay the interest to the bank. Hence, he can perfectly hedge his labour market risk indeed, if the pay-offs are perfectly negatively correlated, and if the expected return in the stock market over and above the cost of capital is equal to the value at risk in the labour market. If he can borrow at all to hedge his labour market risk, he will no doubt have to pay a risk premium over the safe investment interest rate. This risk premium then, he will have to recover in the labour market.

7. DESIGNING SCHOOL STRUCTURES AND CURRICULA

School structures differ widely between countries in terms of horizontal and vertical differentiation, that is, by type of ability and interest and by level of ability. Within schools, there is often a wide variety of curricula, often grouped into separate tracks. One may reflect on these differences, both in terms of optimal individual choices for given educational structures or as the question for optimal design of school structures and curricula. There appears to be no (economic) literature at all on these questions. Some analysis is given in Hartog (1991), but it would be most relevant to develop that approach in an environment of uncertainty. Again, the analogy with the finance literature is interesting. Investment in education is not investment in a good with a single quality, as if buying a single security, but it is rather like buying a portfolio, made up by the different courses that are taken. As noted, one may analyse this as optimal portfolio choice from the demand side, for a given individual but also as the optimal supply, by school administrators or legislators. A starting point may be Rosen (1983) on skill specialisation and extend it with risk.

8. CONCLUSION: A RESEARCH AGENDA

The reflections above can be used to distil an agenda for future research. The order is not by priority, but rather by analytical structure.

1 Ex ante versus ex post returns.

Ex post, the rate of return measures how much individuals benefited from their education, in comparison with their best non-chosen alternative. The non-chosen alternative cannot be observed, which brings in the issues of self-selection. Ex post, the return is measured for individuals who attained some level of education, for particular exits from the education system (for obvious reasons measured in cross-sections rather than full lifetime panel data). The ex ante return takes into account that success in the educational program is not guaranteed. It includes the option values of next moves in the education system. The ex ante return can be estimated by considering the returns for all those who once entered some
education, as in some reduced form aggregate, or one may distinguish the potential exits separately, as in Altonji (1993). Hence, ex-post is for exits, ex ante is for entries. Ex ante returns require earnings data on individuals who started some education. In this way, one may calculate ex ante returns for every next step, or every next year spent in the schooling system. Such calculations require earnings data for a cohort of individuals all entering a particular education, and then proceed to different exits.

2. characterise variance and risk in the return to education.

We can characterise the variance in the ex post rate of return to education, by considering earnings differentials over the entire lifecycle, and relate these to the amount that has been invested. We can use the formula cited by Becker, and collect the required information on the variance of cost, of returns, their covariance and on the inter-temporal covariance of earnings. The covariance between cost and returns points to the relevance of comparative advantage and the Roy model.

Characterising ex ante risk of the investment in education is more complicated, as it also requires taking into account the ex ante probability distribution of routes through the education system (and of potential exits).

As far as I know, there have been no calculations of the variance of the rate of return or the ex ante risk of investment in schooling, other than the illustrative calculations by Becker.

3. test demand for education in relation to risk

There are a few analyses of human capital investment under risk. They are all based on the EU hypothesis. Empirically, there are only a few tests of the core result that increasing risk reduces demand for education. Empirical work requires a measure of risk and variation in it. Cross-section data seem most adequate, with earnings variance conditional on education as the measure of risk. Note that the proper measure of risk is ex ante risk, which complicates estimations. One would also like information on individual risk attitudes, to allow for measuring self-selection. The theoretical models can be extended with alternatives to EU, such as CPT and the very straightforward MV approach.

It’s probably best to consider models for particular educations (types of schools) rather than models with continuous investment in human capital. The analyses can be extended to include details of schooling choices such as curriculum and type of school (eg vocational versus academic).

4. test for risk compensation in wages

Given the formulas derived above, testing compensation for failure in education is quite simple: for an earnings function that includes completed levels of education, just add the probability of failing those education levels (i.e., add failure rates to an ex post rate of return estimate).

The structural model in Hartog-Vijverberg was based on EU. An alternative would be to consider CPT. Using M/V probably generates a relatively simple model: adding variance and
skewness to an earnings function would be sufficient to estimate the coefficients of the indifference curve. However, the measures should relate to ex ante lifetime variability.

Analysis in a broader context is probably not needed. It must be considered highly implausible that individuals are indifferent to risk in the labour market because they would be able to hedge it away at zero cost. As long as costless hedging (insurance) is not feasible, it is legitimate to test for compensation in wages. Nevertheless, it is interesting to estimate the covariance between capital and labour market returns.

An interesting alternative to compensation in wages for the risk born by workers (a positive effect) is the compensation for the risk born by employers, according to internal labour market theories such as that by Harris and Holmstrom (1982). According to these theories, employers insure workers against unpleasant surprises in revealing their true abilities as they accumulate work experience. In return for not lowering their wage when they turn out to be less productive than average, employers reduce the starting wage and gradually reduce this premium as information becomes more precise. This would imply that experience-earnings profiles are flatter when there is more uncertainty about true abilities. This can be tested, provided one either has a direct measure on ability or output uncertainty or the truncation of the distribution observed under insurance is taken into account. The results in Hartog and Vijverberg seem to refute the employer insurance thesis, but admittedly, we did not consider truncated slope effects.

5. Analyse optimal school structure

As schooling is an investment for a very long time-span, much longer than common financial or business investments, its optimal construction to be profitable in all seasons, is an important issue. Becker alluded to these issues in Human capital, p 204, suggesting that a liberal arts education would provide the flexible investment that would be recommended from very long pay-off periods. One may analyse (socially) optimal curriculum structures in terms of desirable risk properties, i.e. considering school structures as an optimum portfolio. Using the Mean-Variance approach, it is easily shown that his conclusions on the dominance of specialisation when the cost function is separable, no longer holds under conditions of risk.
Appendix. MV or EU?

A common model to analyse choice under uncertainty is the model of expected utility maximization, with a utility function that is concave in income, the EU model. An alternative modelling strategy is to specify a utility function in mean and dispersion of income, \( V(\mu, \sigma) \), the MS, or mean-standard deviation model, widely used in financial economics. Intuitively, one expects that under conditions, these approaches are equivalent. Meyer (1987) specifies the conditions and formalizes the intuition.

Meyer considers a choice set in which all random variables \( Y_i \) only differ by location and scale parameters, the LS condition. Formally, two cumulative distribution functions \( G_1(.) \) and \( G_2(.) \) are said to differ only by location parameters \( \alpha \) and \( \beta \) if \( G_1(x) = G_2(\alpha + \beta \cdot x) \) with \( \beta > 0 \). Then, with \( X \) defined as the normalized variable \( X = (Y_i - \mu_i)/\sigma_i \) with \( \mu_i \) and \( \sigma_i \) mean and standard deviation of \( Y_i \), all \( Y_i \) are equal in distribution to \( \mu_i + \sigma_i \cdot X \). Expected utility from \( Y_i \), for an individual with utility function \( u(.) \), can then be written as

\[
EU(Y_i) = \int_{\mu_i}^{\mu_i + \sigma_i} u(x) \cdot dF(x) \equiv V(\sigma_i, \mu_i)
\]

Meyer proves seven properties for the relation between the functions \( V(.) \) and \( u(.) \). \( \Pi(\sigma, \mu) = -V_\sigma / V_\mu \) is the slope of an indifference curve in the MS model.

1. \( V_{\mu}(\sigma, \mu) \geq 0 \) for all \( \mu \) and all \( \sigma \geq 0 \) if and only if \( u'(\mu + \sigma x) \geq 0 \) for all \( \mu + \sigma x \).

2. \( V_{\sigma}(\sigma, \mu) \leq 0 \) for all \( \mu \) and all \( \sigma \geq 0 \) if and only if \( u''(\mu + \sigma x) \leq 0 \) for all \( \mu + \sigma x \).

3. \( \Pi(\sigma, \mu) \geq 0 \) for all \( \mu \) and all \( \sigma > 0 \) if \( u'(\mu + \sigma x) \geq 0 \) and \( u''(\mu + \sigma x) \leq 0 \) for all \( \mu + \sigma x \).

4. \( V(\sigma, \mu) \) is a concave function for all \( \mu \) and all \( \sigma > 0 \) if and only if \( u''(\mu + \sigma x) \leq 0 \) for all \( \mu + \sigma x \).

5. \( \partial \Pi(\sigma, \mu) / \partial \mu \leq (=, \geq) 0 \) for all \( \mu \) and all \( \sigma > 0 \) if and only if \( u(\mu + \sigma x) \) displays decreasing (constant, increasing) absolute risk aversion for all \( \mu + \sigma x \).

6. \( \partial \Pi(\sigma, \mu) / \partial \sigma \geq (=, \leq) 0 \) if and only if \( u(\mu + \sigma x) \) displays increasing (constant, decreasing) relative risk aversion for all \( \mu + \sigma x \).

7. \( \Pi_1(\sigma, \mu) \geq \Pi_2(\sigma, \mu) \) for all \( (\sigma, \mu) \) if and only if \( u_i(\mu + \sigma x) \) is more risk averse than \( u_2(\mu + \sigma x) \) for all \( \mu + \sigma x \).

Thus, if random variables affected by individuals' choice only differ in location and scale, we can represent the preferences under expected utility maximization with a concave utility function.
function by a preference relation in $\mu$ and $\sigma$ that generates the same preference ordering.\footnote{A direct application of property $\mu$ relates to the specification of $\nu(\sigma, \mu)$ by Low and Ormiston (19..), where $\Pi_{\mu} > 0$ implies increasing absolute risk aversion.}

The MV model usually is a simpler representation of preferences, and hence, usually simplifies analysis. If an economic model involves a random variable that is affected by choice only in location and scale, we can simply solve the model by maximizing $V(\sigma, \mu)$ subject to the restrictions on $\sigma$ and $\mu$. A simple application is the schooling choice model. Suppose, the capital market is perfect, and the lifetime choices can be separated in decisions on lifetime income and lifetime intertemporal allocation of consumption. With $\mu(s)$ and $\sigma(s)$ reflecting the dependence of $\mu$ and $\sigma$ of discounted lifetime income distribution on length of schooling $s$, maximizing $V(\sigma(s), \mu(s))$ requires

\[
\frac{\partial \mu(s)}{\partial s} = -\frac{V_\sigma}{V_\mu} \frac{\partial \sigma(s)}{\partial s} = \Pi \frac{\partial \sigma(s)}{\partial s} \quad \alpha \quad \frac{\partial \mu / \partial s}{\partial \sigma / \partial s} = \Pi
\]

A risk neutral individual has $V_\sigma = \Pi = 0$ and hence, will choose $s$ to maximize lifetime mean earnings. Suppose, expected discounted lifetime earnings $\mu(s)$ is parabolic in schooling $s$. Then, a risk averse individual ($\Pi > 0$) will choose less schooling and a risk loving individual ($\Pi < 0$) will choose more schooling than a risk neutral individual if dispersion of lifetime discounted earnings increases in schooling, and the converse holds if dispersion is decreasing in schooling.
REFERENCES


Davis, S. J. and P. Willen (2000), Risky labor income and portfolio choice, University of Chicago and Princeton Working Paper (available on Davis’ website, Chicago GSB)


Harmon, C., V. Hogan and I. Walker (2001), Dispersion in economic return to schooling, CEPR Discussion Paper 3037


Hartog, J. (1991), Capabilities, allocation and earnings, Boston: Kluwer

Hartog, J., A. Ferrer-i-Carbonell and N. Jonker (2002), Linking measured risk aversion to individual characteristics, Kyklos 55(1), 3-26


Jacobs, B. (1998), A closed form solution for the human capital model with risk, working paper SCHOLAR, Amsterdam


Powel, Schubert and Gysler (2001), Discussion Paper, Schweizerische Technische Hochschule


Wakker, P. (1994), Separating marginal utility and probabilistic risk aversion, Theory and Decision, 36 (1), 1-44

Weiss, Y. (1972), The risk element in occupational and educational choices, Journal of Political Economy, 80 (6), 1203-1213