

IZA DP No. 6166

**It's the Opportunity Cost, Stupid!
How Self-Employment Responds to Financial
Incentives of Return, Risk and Skew**

Peter Berkhout
Joop Hartog
Mirjam van Praag

November 2011

It's the Opportunity Cost, Stupid! How Self-Employment Responds to Financial Incentives of Return, Risk and Skew

Peter Berkhout

RIGO Research Institute

Joop Hartog

*University of Amsterdam,
Tinbergen Institute and IZA*

Mirjam van Praag

*University of Amsterdam,
Tinbergen Institute and IZA*

Discussion Paper No. 6166
November 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

It's the Opportunity Cost, Stupid! How Self-Employment Responds to Financial Incentives of Return, Risk and Skew^{*}

There is no robust empirical support for the effect of financial incentives on the decision to work in self-employment rather than as a wage earner. In the literature, this is seen as a puzzle. We offer a focus on the opportunity cost, i.e. the wages given up as an employee. Information on income from self-employment is of inferior quality and this is not just a problem for the outside researcher, it is an imminent problem of the individual considering self-employment. We also argue that it is not only the location of an income distribution that matters and that dispersion and (a)symmetry should not be ignored. We predict that higher mean, lower variance and higher skew in the wage distribution in a particular employment segment reduce the inclination to prefer self-employment above employee status. Using a sample of 56,000 recent graduates from a Dutch college or university, grouped in approximately 120 labor market segments, we find significant support for these propositions. The results survive various robustness checks on specifications and assumptions.

JEL Classification: J24, L26

Keywords: entrepreneurship, self-employment, wage-employment, income distribution, income risk, income skew, income variance, occupational choice, labor market entry, labor market segments, opportunity cost

Corresponding author:

Mirjam van Praag
Amsterdam Center for Entrepreneurship (ACE)
Faculty of Economics and Business (FEB)
Roetersstraat 11
1018 WB Amsterdam
The Netherlands
E-mail: c.m.vanpraag@uva.nl

^{*} Excellent and very useful comments were received from Monique de Haan, Thomas Astebro and Simon Parker.

1. Introduction

Imagine someone just graduating from school and considering whether to work as an employee or in self-employment. Economists perceive that differences in income will be among the factors determining the choice. But obtaining information on potential income from self-employment is a formidable job, much more so than obtaining information on potential pay as an employee. There are far fewer self-employed workers than employees with a given education to act as informants, accounting rules for exactly determining income from independent business are not unequivocal, the range of possible outcomes is much wider and often not even foreseeable, professional abilities and competences, as determinants of potential income, are harder to predict than performance in a controlled employee status and the environment for a self-employed worker is inherently more dynamic. The literature often stresses that a researcher is unable to obtain good, reliable data on self-employment income for econometric analysis. We argue that this is not just a problem for the outside researcher, but an imminent problem that also confronts the subject of her research.

Effects of financial incentives in the choice between employee or self-employment status have been hard to establish, but given the problems noted above this should in fact be no surprise. There is a marked asymmetry in the quality of information on potential income from self-employment and from employee status. Therefore, in our model of the choice of employment status, we will focus on the opportunity cost of self-employment rather than on the benefits. Indeed, we report robust support for the moments of the wage distribution as determinants of the decision to become self-employed. A lower mean wage income in one's labor market segment increases the probability of choosing for self-employment. Higher variance and lower skew in the distribution of wage incomes also increase it. And as in most of the literature, we find no effect of self-employment income.

The weak and mixed results on the effect of financial incentives in the choice for self-employment have been presented as “something of a puzzle since they suggest that entrepreneurs do not respond robustly to pecuniary incentives... However, there is so much economic evidence that individuals adjust their behavior in response to changes in relative prices that it would be puzzling if the same calculus ceased to apply entirely in the realm of entrepreneurship as an occupational choice.” (Parker, p. 110).¹ Both Astebro (2010) and Parker (2009) stress the information problem as a prominent factor in potential explanations for the puzzle. Parker (2009, p 110) refers to poor data quality and measurement error as common phenomena in observed entrepreneurial earnings levels. In fact, entrepreneurial incomes remain often unreported or, if not, they are systematically underreported, both in tax filings and public surveys. Astebro even refers to lack of interest among the self-employed: ‘Entrepreneurs are wary of revealing accurate income data to third parties’ (Astebro, 2010, p. 36). Both authors note the difficulties in extracting uniform entrepreneurial income measures from the reported data and make them comparable to wage earnings. Some researchers define the

¹ Hyytinen et al. (2008) call this the ‘returns to entrepreneurship puzzle’, see also Astebro (2010) or Hamilton (2000). We are aware of the fact that “entrepreneur” and “self-employed” are not identical concepts. We use “entrepreneur” only when we discuss the literature. When we present our own results, we will refer to self-employment.

entrepreneur's income as the amount the entrepreneur draws from the firm on an annual basis, whereas other researchers use the firm's net profit, or alternatively, the sum of the value growth of the firm and 'draw' as a proxy of the entrepreneur's income (Parker, 2009, p. 363).² ³ No doubt, this lack of a uniform concept of self-employment income is also a problem for individuals who consider self-employment.

Not only researchers, but also the subjects themselves will have a much sharper picture of the opportunity cost of self-employment than of the income it will bring. It is even conceivable they actually care more about opportunity cost than about financial returns, as non-pecuniary benefits are often found to be among the key arguments for preferring self-employment: autonomy (Benz and Frey, 2008), a tendency towards entrepreneurship caused by genetic factors (Nicolaou, Cherkas, Hunkin and Spector, 2008), cognitive biases arising from overoptimism (Lowe and Ziedonis, 2006; Dushnitsky, 2010) and/or overconfidence (Hayward, Shepherd and Griffin, 2006). A focus on financial opportunity cost rather than on financial returns aligns well with this perspective on potentially self-employed workers.

We do not only consider the opportunity cost in terms of mean incomes, but we acknowledge the possibility that utility differences between the two options will, more in general, also result from entirely different probability distributions of incomes offered. No doubt, the three dominant characteristics of a distribution are location, dispersion and (a)symmetry. Higher variance in earnings of entrepreneurs compared to employees is a stylized fact (Astebro, 2010; Parker, 2009) and requires a risk premium for risk averse individuals (King, 1974; Cramer et al., 2002; Caliendo et al., 2009; Bonin et al., 2007). Recent studies have also provided evidence that the distribution of entrepreneurial earnings has stronger (positive) skew than the wage distribution. Skew may play a role for entrepreneurship choices as 'a few extremely high prices have a disproportionately great attractive force', Marshall (1890, 1930, p. 554). Indeed, economic theory has established that declining absolute risk aversion, an almost inevitable hypothesis, requires a positive appreciation of skew (Astebro et al., 2009; Tsiang, 1972; Hartog, 2011). The relevance of skew in choices under uncertainty has been demonstrated in several applications such as gambling and betting (Garrett and Sobel, 1999; Golec and Tamarkin, 1998) or in the literature of lifetime wealth accumulation, where the appreciation of skew is called prudence (Gollier, 2001, p 238). A positive premium for variance and a negative premium for skew in wages across different types of educations, established for several countries in different settings (Hartog, 2011) indicates that variance and skew are relevant for choice of education cum occupation. Not accounting for the second and third moments of the income distributions may flaw the comparison and thus be an explanatory factor in the lack of consistent support of the role of financial incentives.

² An additional complicating factor is that many business owners do not earn 'business incomes' but pay themselves a wage in their incorporated business. If this is the case, the entrepreneur can hardly be distinguished from a wage employee. Entrepreneurs often incorporate their business when they have personnel and thus run larger (and more successful) firms. The difficulty of distinguishing these entrepreneurs from wage employees may thus lead to a smaller sample of entrepreneurs and underestimates of true entrepreneurial earnings.

³ Researchers may also deal incorrectly with the entrepreneurs' negative incomes and with top-coding or may erroneously include 'returns to capital' to their (labor) income measures (Parker, 2009, p. 368).

We derive propositions from a simple model on the role of financial incentives for the choice between self-employment and wage employment. Our observations are individuals in self-employment or wage employment who recently graduated from more than 100 different types of tertiary education. We argue, precisely based on the arguments outlined above, that the individuals are poorly informed on incomes from self-employment and we focus on the effects of moments of the distribution of wage income. We predict that a higher mean wage income increases the opportunity costs of self-employment and thus reduces the probability of entrepreneurship. A higher variance pushes up the inclination to self-employment, as one gives up a more risky alternative. A higher skew, the third moment, reduces this inclination as one would give up a better probability to end up in an extended upper tail of the wage distribution. By simultaneously analyzing the effects of the first three moments of the wage distribution we squarely acknowledge one of the caveats discussed above, the incompleteness of a comparison based on only mean (or median) income levels.

By focusing on opportunity cost as expressed in the wage distribution, we avoid the problem of poor measurement of self-employment income. We do so by estimating our model on data covering 118 labor market segments defined from detailed information about major track in field of study of 56,000 graduates. The data apply to new graduates who enter the labor force for the first time and have not yet been tied up in any of the alternatives. As a consequence our sample is homogeneous in terms of (no) labor market experience and education level.

The validity of our approach hinges on three assumptions. The first assumption is made when defining labor market segments as fields of study with well defined curriculums in vocational colleges or universities leading to a specific degree (e.g sociology, physics, fiscal law, food technology, etc). We postulate that the relevant distribution of wages is defined by individuals with a particular type of education. We thus assume intra-group homogeneity in terms of the earnings distributions facing the individuals (Reich et al., 1973) relative to the wage distribution across labor market segments. We assess the validity of this assumption by also considering a random grouping into labor market segments.

Second, we assume that individuals know the distribution of wages within their labor market segment but do not know the education specific distribution of income from self-employment. Thus, we assume that students only observe a general distribution of earnings generated by self-employment that is common for all labor market segments. The assumption is motivated by the verifiable facts that there are very few entrepreneurs within each labor market segment and that their incomes are less observable. In our robustness checks, we use as much information on self-employment income as our data allow. But even if the assumption were violated, this would be econometrically harmless as long as the moments of the distributions of wages and of self-employment incomes are independent.⁴ We find no evidence of strong correlations between any of the three moments of the two income distributions; our sensitivity checks are compatible with (modest) underestimation of effects of opportunity cost.

⁴ If the financial attractiveness of the entrepreneurship option were positively correlated with the financial attractiveness of the alternative, this would lead to a downward bias of the true effect. On the contrary, if there were a negative correlation between these two distributions, the effect of pecuniary return, risk and skew in wage employment on entrepreneurship choices would be overestimated.

The third assumption is that there are no confounding factors underlying the choice for a particular field of study that affect both its wage distribution and the likelihood of self-employment. We discuss the role of risk attitude, ability and immanent distribution of productivity by education and we claim that our results survive concerns on these grounds.

The remainder of the paper is organized as follows. Section 2 discusses the theoretical model and the resulting propositions. In Section 3 we discuss the data. In Section 4 we present results of testing the propositions and assumptions. Section 5 concludes.

2. Theoretical model and propositions

A graduate entering working life, after completing an education, has a choice between self-employment and working for a wage as an employee. Both options bring financial returns (w as wage earner, p as self-employed) and non-pecuniary returns (ψ). Non-pecuniary returns reflect the individual's appreciation of the relevant characteristics of self-employment versus employee work such as autonomy (Hyytinen et al., 2008), the differential scope of activities and tasks (Benz and Frey, 2008), the greater insecurity in other dimensions than income (Hamilton, 2000) and a different social status (Parker and Van Praag, 2010).

We assume that the utility derived from working, either as a self-employed entrepreneur (SE) or as a wage employee (WW), is linearly separable in the utility derived from w or p and ψ :

$$U_{ww} = U(w) + U(\varphi_{ww}) \quad (1a)$$

$$U_{se} = U(p) + U(\varphi_{se}) \quad (1b)$$

Assuming separability implies that the valuation of these non-pecuniary aspects is independent of income. Graduates base their occupational choice on a comparison of expected utility levels in each of the occupations. The likelihood of choosing self-employment can thus be derived from:

$$P(SE) = P[E(U_{se}) > E(U_{ww})] \quad (2)$$

Substituting (1) in (2) and expressing the differences in expected utility derived from non-pecuniary factors in both occupations within the same segment⁵ by a (latent) individual specific variable $\varepsilon = E(U(\varphi_{ww})) - E(U(\varphi_{se}))$, it follows that

$$P(SE) = P[\varepsilon < E(U(p)) - E(U(w))] \quad (3)$$

⁵ Segments are defined by education, see below.

where \mathcal{E} remains unobserved and reflects the individual-specific difference in non-pecuniary returns between self-employment and wage-employment within the given labor market segment. Suppose the wage distribution (w) and entrepreneurial income distribution (p) are characterized by the following parameters:

$$E[w] = \mu_w; \quad E[(w - \mu_w)^2] = \sigma_w^2; \quad E[(w - \mu_w)^3] = \kappa_w^3 \quad (4)$$

$$E[p] = \mu_p; \quad E[(p - \mu_p)^2] = \sigma_p^2; \quad E[(p - \mu_p)^3] = \kappa_p^3 \quad (5)$$

Using Taylor series approximation, we derive expressions for the expected utility of wage-employment versus self-employment in equations (6) and (7), respectively, where $U(\mu_w)$ is expanded around \bar{y} (the expected income level for graduates, i.e., a weighed average of μ_w and μ_p)

$$\begin{aligned} E[U(w)] &= E[U(\mu_w) + U'(\mu_w)(w - \mu_w) + \frac{1}{2}U''(\mu_w)(w - \mu_w)^2 + \frac{1}{6}U'''(\mu_w)(w - \mu_w)^3 + \dots] \\ &= U(\bar{y}) + U'(\bar{y})(\mu_w - \bar{y}) + \sum_{i=2}^{\infty} \frac{1}{i!} U^{(i)}(\bar{y})(\mu_w^i - \bar{y}) + \frac{1}{2}U''(\mu_w)\sigma_w^2 + \frac{1}{6}U'''(\mu_w)\kappa_w^3 + \dots \end{aligned} \quad (6)$$

$$\begin{aligned} E[U(p)] &= E[U(\mu_p) + U'(\mu_p)(p - \mu_p) + \frac{1}{2}U''(\mu_p)(p - \mu_p)^2 + \frac{1}{6}U'''(\mu_p)(p - \mu_p)^3 + \dots] \\ &= U(\bar{y}) + U'(\bar{y})(\mu_p - \bar{y}) + \sum_{i=2}^{\infty} \frac{1}{i!} U^{(i)}(\bar{y})(\mu_p^i - \bar{y}) + \frac{1}{2}U''(\mu_p)\sigma_p^2 + \frac{1}{6}U'''(\mu_p)\kappa_p^3 + \dots \end{aligned}$$

(7)

Substituting these expressions in (3), we obtain:

$$P(SE) = P[\varepsilon_i < a + U'(\bar{y})(\mu_p - \mu_w) + \frac{1}{2}U''(\mu_p)\sigma_p^2 + \frac{1}{6}U'''(\mu_p)\kappa_p^3 - \frac{1}{2}U''(\mu_w)\sigma_w^2 - \frac{1}{6}U'''(\mu_w)\kappa_w^3] \quad (8)$$

where the constant term a absorbs all that has been left out. Equation (8) expresses a standard probit (or logit) model for the choice of self-employment status vis-a-vis the alternative of wage employment. The coefficients can be estimated if we observe sufficient numbers of sufficiently distinct observations for the distribution of wages (w) and self-employment incomes (p).

Employment segments are defined by field of education, as we take individuals' education as completed. This does not rule out potential substitution with graduates from other fields: the effects are simply included in the observed wage distribution. Presumably, individuals consider their perspective over some time horizon.⁶ We assume that individuals, when deciding on their employment status, know the parameters of the wage distribution but do not know where in the wage distribution they would end up if

⁶ An implicit assumption of our approach is that the distribution of wages among recent graduates adequately characterizes the distribution over that horizon. This may be justified by a high correlation of distribution parameters for different experience lengths and by the fact that the individual may still be unaware of the duration of his commitment.

choosing wage employment. We take our observations from 118 distinct wage distributions ($k = 118$ labor market segments defined by a degree field).⁷

For the reasons outlined in the introduction, graduates have far less information on self-employment incomes, let alone self-employment incomes specific to their field of education. We will assume they will use the parameters for self-employment income without differentiating by field of education. In an analysis of self-employment by field of education, the self-employment income prospects are then identical for all individuals and will reduce to a constant; thus, attention fully falls on the parameters of the opportunity cost.

There is no need to deny that originally, this assumption arose out of necessity: with often less than 10 self-employed workers per field one cannot reliably characterize a distribution. But upon reflection, it is not as extreme as it may perhaps seem at first sight. We do not compare highly educated professionals to the local grocery store owner or to a self-employed carpenter: all our respondents have completed tertiary education. And it is actually quite conceivable that subjects indeed use much cruder information on self-employment income than on employee income. To them, just as to us, information on self-employment income is limited by data and definition problems and has often low reliability. These problems are aggravated if one desires information by field of education, with very few examples around to assess income from self-employment. Perhaps, subjects even care little about precise information as other aspects of self-employment matter much more to them. In our robustness tests, we check the validity of the assumption to the extent we can: we include self-employment income parameters by field of education for a subsample with sufficient numbers of observed self-employed workers and we also estimate with self-employment income parameters for aggregates of intrinsically related fields of education. Our core conclusions easily survive. We may note that, statistically, the only requirement is independence: if the variation in the distribution of self-employment income is independent of the variation in the distribution of wages, our estimates are unbiased.⁸

Under this assumption we can estimate the parameters b , c and d by means of the following probit-equation, where a absorbs the effect of the common income distribution for entrepreneurs:

$$P(SE) = P[\varepsilon_i < a - b\mu_w + c\sigma_w^2 - d\kappa_w^3] \quad (9)$$

The coefficients b , c and d are defined in the first, second and third derivative of the utility function; we take them here as positive numbers. With utility increasing in the first moment (from a taste for consumption), the second derivative of utility negative (from risk aversion) and the third derivative positive (from skew affection, as implied by declining absolute risk aversion, cf Tsang, 1972), we predict that the probability of self-employment among graduates from a field of education will decrease in the mean of the wage

⁷ In the Netherlands, tertiary education is not characterized by a major, but it is a field defined by a specific curriculum right from the beginning, e.g. economics or chemistry.

⁸ This holds conditional on ignoring higher moments in our Taylor series expansion of the utility function.

distribution in that field, increase in the variance of wages and decrease in the skew of wages for the graduate's field of education.⁹

3. Data

In this section, we shall first discuss briefly the relevant features of the Dutch educational system. We will then discuss the data sources and descriptive statistics.

3.1 *The Dutch system of higher education*

In the Netherlands, 52 vocational colleges and 14 universities offer programs in higher education. Both types of post-secondary education offer a broad array of study programs at the Bachelor level. During the time period captured by our data, universities offered only combined four-year programs leading to a Master's degree, whereas the vocational colleges offer more practically oriented bachelor programs of four years.¹⁰ Students starting these programs were typically 18 years old and fresh from high school. The number of students in the academic year 2008/2009 who graduated from vocational colleges with a bachelor degree was 52,000, whereas the number of students that entered the labor market with a Master degree was approximately 28,000.

3.2 *Data sources and definitions*

Our sample consists of 79,415 recent graduates from tertiary education. It is a random draw of approximately 7,500 recent graduates per annum in the years 1999 to 2008, from a yearly survey to monitor the labor market outcomes of graduates in the 118 largest degree fields, half in higher vocational education and half in university education fields. The special feature of the monitor is the detailed information on study background and early career earnings of many individuals in many degree fields in tertiary education. Individuals fill out extensive questionnaires about their study program and grades, the activities they undertook as a student, individual background and current labor market situation, such as job search activities, occupational status (unemployed, self-employed, wage-employed) and details about their income. They do so, on average, 20 months after graduation, in January. Thus, our sample includes graduates from the academic years 1997/1998 until 2006/2007.

The Dutch research institute SEO has been commissioned by the prominent weekly magazine

⁹ We have abstracted from the fact that the derivatives and hence the coefficients will vary with μ_w . Presumably, this is a small, secondary, effect.

¹⁰ The bachelor master (BAMA) structure was introduced in the Dutch system of higher education in 2002-2003. The first graduations within the new structure took place in 2005-2006 (academic bachelor program of 3 years) and 2006-2007. Only the last wave of observations (January, 2008) may theoretically include these university bachelor graduates since the 2008 wave is held among graduates of the academic year 2005-2006. The few students who entered the labor market with an academic bachelor degree in the first possible year, i.e., 2005-2006 are not included in the 2008 sample. Therefore, the distinction between a degree from a vocational college and a university is implied by the distinction between bachelor and masters degrees.

Elsevier to gather and analyze these data. *Elsevier* publishes a leading special issue every year about the labor market prospects of the 118 specific degree fields (80% of all graduates from tertiary education have degrees from these fields).¹¹ Appendix Tables A1 and A2 list all vocational and academic degree fields respectively (first column) and the number of wage-employed and self-employed observations in each (second column).

Our dependent variable is occupational status. It is based on the answer to the question: “What is your status in the labor market at the moment?” We are interested in particular in the distinction between self-employed entrepreneurship and wage employment. Table 1 shows how the distribution of individuals among the various answering categories (and non-response, small sized labor market segments or age) diminishes our effective sample size from 79,415 to 56,138 graduates. Respondents could select only one, the most applicable category.

The second key variable is the income distribution per labor market segment. To estimate equation (9), we need observations on the three moments μ_w , σ_w^2 and κ_w^3 . Calculated moments of the wage distribution per degree field are based on the wages of employed respondents in the field who report their income.¹² Income data for the various years are expressed in real terms (for the year 2008) by correcting for inflation (consumer price index). Hourly wages are calculated as monthly income divided by the reported number of hours worked per month. Hourly wages below 5 euro have been set to missing, given the minimum wage laws that apply. Wages higher than three standard deviations above the average wage level in their degree field have also been set as missing. Deleting outliers is relevant because the moments of the wage distribution are sensitive to extreme values.¹³

Table 1
Definition of the sample

| | |
|-------------------------------------|--------|
| Initial sample size | 79,415 |
| Inactive in the labor market | 12,335 |
| Temporary workers | 3,083 |
| Education degree field unknown | 342 |
| Education degree field small (n<50) | 343 |
| Graduate age >30 | 7,175 |
| Effective sample size | 56,138 |

3.3 Descriptive statistics

As shown in Table 2, the average (median) number of individuals in our sample with identical degrees is 476 (465). This defines the average (median) size of the 118 labor market segments. The largest labor market segment includes 1502 individuals (economics, Msc), the smallest only 59 (horticulture and agriculture, Bsc).

¹¹ The data have been obtained with the permission of Elsevier and SEO Economic Research.

¹² Tables A1 and A2 show the wage moments and the numbers of observations per degree field with a wage income.

¹³ Hourly self-employed entrepreneurial incomes are calculated likewise, although ‘outliers’ are not excluded. For self-employed entrepreneurs, minimum wages do not apply and occasional very high incomes may occur. Therefore, and also because of the low number of observations we start from, we do not delete extreme values in the income distribution of entrepreneurs. All results remain qualitatively the same when we delete these observations from the self-employed income distribution (or, alternatively, when we do not delete extreme observations from the wage income distributions).

Only 3.6 percent of labor market participants are self-employed.¹⁴ It is a common finding that the self-employment rate among recent graduates is very low (Parker, 2009). Table 2 also shows that the fraction of entrepreneurs varies considerably per degree field. The minimum is zero. This applies to 3 out of 118 degree fields: “Food technology” (Bachelor), “Special needs teacher” (Bachelor) and “Language and culture” (Master). The three degree fields with the highest fraction of entrepreneurs are “Visual arts & design” (Bachelor) with a fraction of 35%, “Theater” (Bachelor, 20%) and “Dance” (Bachelor, 19%).¹⁵ These education degrees are all offered in the vocational colleges. The degree fields in universities with the highest fractions of entrepreneurs are “Film, television and theater studies” (13%) and “Industrial design” (12%).

Table 2
Descriptive statistics: Labor market segments, occupational choice and sample sizes

| | Individuals (n) | Per degree field (k =118) | | | | |
|-------------------------------|-----------------|---------------------------|------------|----------------|------------|------------|
| | | <i>Average</i> | <i>Med</i> | <i>Std dev</i> | <i>Min</i> | <i>Max</i> |
| # labor market participants | 56,138 | 476 | 465 | 274 | 59 | 1502 |
| # wage earners, income known | 51,080 | 437 | 424 | 259 | 33 | 1418 |
| # self-employed, income known | 1,538 | 13 | 9 | 18 | 0 | 152 |
| % self-employed | 3.64 | 4.3 | 2.6 | 5.8 | 0 | 37 |
| % wage earners, income known | 94.4% | | | | | |
| % self-employed, income known | 75.2% | | | | | |

Only 75 percent of the self-employed report earnings, a clear illustration of our claim in the introduction that ‘Entrepreneurs are wary of revealing accurate income data to third parties’ (Astebro, p. 36, 2010). The combination of a much lower number of self-employed than wage earners per segment and the lower response rates to questions about earnings among self-employed induces us to use the self-employment income data with great caution only. The average (median) number of self-employed per labor market segment who report their income is only 13 (9). Moreover, only 50 out of the 118 labor market segments include at least ten observed incomes of self-employed. The number of labor market segments with at least 20 observed incomes for self-employed is only 19.

Table 3 shows the descriptive statistics of the wage distributions across individuals (first column) and labor market segments (right hand side columns) in terms of the first three central moments, i.e. μ_w , σ_w^2 and κ_w^3 from equation 9. The table shows that labor market segments are not terribly different in terms of the average wage income, ranging from 8.54 to 13.13 per hour with a standard deviation of 0.837. However, the variance of the wage incomes within a labor market segments as well as the skew vary widely across segments. In particular, the standard deviation of κ_w^3 is very large. The wage variance and skew in the population of wage earners are larger than the average variance and skew within labor market segments.

¹⁴ The percentage of entrepreneurs in the sample of individuals differs from the percentage of entrepreneurs in the sample of labor market segments, due to the fact that labor market segments vary in size leading to a different weighing of individual observations.

¹⁵ The degree field with the highest fraction of entrepreneurs is also the one with the highest number of entrepreneurs, i.e. 240.

The lower half of Table 3 shows the (available) statistics of the distribution of entrepreneurial income, p . The usual comparison to the wage distribution applies. Self-employment incomes are somewhat higher on average (11.2 versus 10.2) with a lower median level (9.5 versus 9.8). The variance is more than 12 times larger than the wage variance for employees (77 versus 6). The skew of the p -distribution is almost 100 times the skew of w (3260 versus 34). Consistent with the picture painted by the upper half of Table 3, the variance and skew of p , like for w , are particularly large across labor market segments, more so than within.

Table 3
Descriptive statistics of the wage and self-employment income distributions

| Income statistics in €/h | n _w = 51,080 wage observations, incomes known | | k = 118 Labor market segments/degree fields | | | |
|--|--|--|---|---------|---------|--|
| | Mean | Mean | Std dev | Min | Max | |
| μ_w | 10.244 | 10.081 | 0.837 | 8.540 | 13.128 | |
| σ_w^2 | 6.017 | 5.168 | 3.801 | .886 | 32.352 | |
| κ_w^3 | 33.683 | 25.173 | 57.441 | -.0449 | 450.99 | |
| Median of w | 9.848 | 9.975 | | | | |
| Self-employed, incomes known, (n _p = 1,538) | | Labor market segments/degree fields (k = 50, includes segments ≥ 10 observations of p) | | | | |
| μ_p | 11.241 | 11.471 | 2.357 | 7.136 | 18.404 | |
| σ_p^2 | 77.194 | 62.744 | 82.109 | 9.730 | 411.670 | |
| κ_p^3 | 3260 | 2172 | 5934 | -15.066 | 29144 | |
| Median of p | 9.490 | 10.550 | | | | |

*The statistics for wage income earners, when based on the same selection of 50 labor market segments as the sample for which the self-employment income statistics are shown, are similar to those for $k=118$.

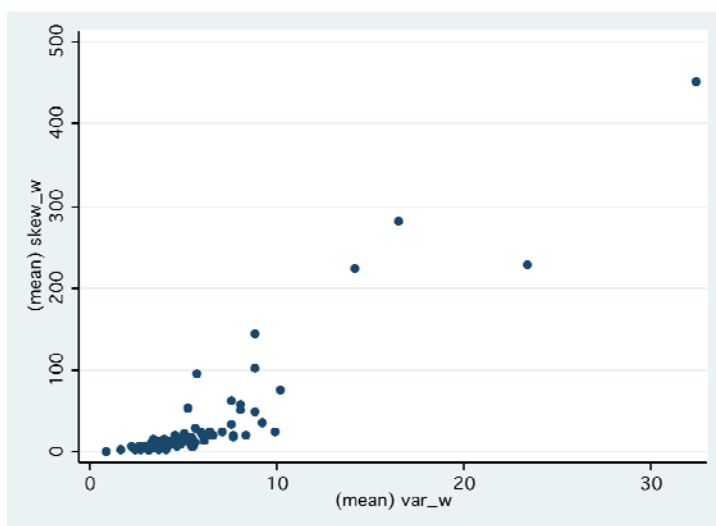
The upper left quadrant of Table 4 depicts the correlation levels between the first three central moments of the wage distribution of each of the labor market segments. These correlation levels are all positive. The correlation between σ_w^2 and κ_w^3 amounts to 0.92 and points to problems of multicollinearity. At that level we cannot trust estimates of separate effects of variance and skew at face value, especially if the standard errors are large and the coefficient estimates change by much in response to small changes in the model (in terms of adding/excluding either variables or observations). We therefore consider not only the second and third moments of the wage distribution but in an alternative set of specifications replace them by the (standardized) coefficient of variation (σ_w/μ_w) and the commonly used (scale free) measure of skew κ_w^3/σ_w^3 , respectively. The lower left quadrant of Table 4 shows that using these scale free measures of risk and skew reduces the three relevant correlations to the manageable levels of 0.29, -0.04 and 0.56 respectively.

We pursue a second way of alleviating the potential multicollinearity problem by a visual inspection of the 118 data-points in the space defined by σ_w^2 (on the horizontal axis) and κ_w^3 (on the vertical axis), see Figure 1.

| | k=118 labor market segments | | | | k=114 labor market segments | | | |
|-------------------------|-----------------------------|--------------|--------------|------------------|-----------------------------|--------------|--------------|------------------|
| | μ_w | σ_w^2 | κ_w^3 | σ_w/μ_w | μ_w | σ_w^2 | κ_w^3 | σ_w/μ_w |
| μ_w | 1.000 | | | | 1.000 | | | |
| σ_w^2 | 0.543 | 1.000 | | | 0.433 | 1.000 | | |
| κ_w^3 | 0.384 | 0.919 | 1.000 | | 0.071 | 0.689 | 1.000 | |
| σ_w/μ_w | 0.294 | 0.894 | 0.798 | 1.000 | 0.061 | 0.909 | 0.702 | 1.000 |
| κ_w^3/σ_w^3 | -0.036 | 0.401 | 0.573 | 0.557 | -0.156 | 0.387 | 0.857 | 0.506 |

Figure 1

Descriptive statistic: the relationship between σ_w^2 and κ_w^3 for $k = 118$ labor market segments



The figure indicates that excluding the four labor market segments in the upper right part of the graph may decrease the correlation between risk and skew substantially. The four labor market segments that are excluded –from right to left- are ‘physiotherapy (Bachelor)’, ‘music (Bachelor)’, ‘logopedy’ (Bachelor), and ‘pedagogy’ (Master). This decreases the correlation between σ_w^2 and κ_w^3 successively from 0.92 to 0.85, 0.81, 0.75 to, finally, 0.68 for $k = 114$. The right hand side of Table 4 shows all resulting correlations. In the sequel, we consider four sets of analyses; with scaled versus unscaled measures of risk and skew and with $k = 118$ versus $k = 114$. This will reduce the impact of multicollinearity problems and show the robustness of the results to changes in the setup of the measures and the sample.

Table 5 shows the descriptive statistics of the control variables that are included in the regression equations, either as a further test of robustness or as a test of assumptions. The factors that are commonly used to explain the variation in self-employment choices in a sample of labor market participants are included (Parker, 2009). These are labor market characteristics (year and region), personal characteristics (gender, age,

parental education levels and risk attitude) and human capital variables such as education levels and grades.¹⁶

Table 5
Descriptive statistics of the control variables

| | All individuals (n=56,138) | | Wage employees (n _w =54,092) | | Self-employed (n _{pp} =2,046) | |
|--|-------------------------------|---------------|--|---------------|---|---------------|
| | <i>Mean</i> | <i>St Dev</i> | <i>Mean</i> | <i>St Dev</i> | <i>Mean</i> | <i>St Dev</i> |
| Labor market characteristics | | | | | | |
| Year = 1999 (dummy) | 0.110 | 0.313 | 0.110 | 0.313 | 0.115 | 0.319 |
| Year = 2000 (dummy) | 0.096 | 0.295 | 0.095 | 0.293 | 0.099 | 0.299 |
| Year = 2001 (dummy) | 0.090 | 0.286 | 0.090 | 0.286 | 0.083 | 0.276 |
| Year = 2002 (dummy) | 0.086 | 0.280 | 0.086 | 0.280 | 0.077 | 0.267 |
| Year = 2003 (dummy) | 0.100 | 0.300 | 0.102 | 0.303 | 0.062 | 0.241 |
| Year = 2004 (dummy) | 0.103 | 0.304 | 0.104 | 0.305 | 0.080 | 0.271 |
| Year = 2005 (dummy) | 0.118 | 0.323 | 0.118 | 0.323 | 0.126 | 0.332 |
| Year = 2006 (dummy) | 0.106 | 0.308 | 0.106 | 0.308 | 0.121 | 0.326 |
| Year = 2007 (dummy) | 0.112 | 0.315 | 0.110 | 0.313 | 0.139 | 0.346 |
| Year = 2008 (dummy) | 0.079 | 0.270 | 0.079 | 0.270 | 0.099 | 0.299 |
| Region South of NL (dummy) | 0.205 | 0.404 | 0.206 | 0.404 | 0.174 | 0.379 |
| Region North of NL (dummy) | 0.076 | 0.265 | 0.076 | 0.265 | 0.073 | 0.260 |
| Region West of NL (dummy) | 0.529 | 0.499 | 0.526 | 0.499 | 0.578 | 0.494 |
| Region East of NL (dummy) | 0.190 | 0.392 | 0.190 | 0.392 | 0.175 | 0.380 |
| Personal characteristics | | | | | | |
| Male (dummy) | 0.447 | 0.497 | 0.444 | 0.497 | 0.514 | 0.500 |
| Age (years effective range 21-30) | 26.025 | 1.729 | 26.012 | 1.724 | 26.388 | 1.834 |
| Parents' education level (normalized average) | 0.000 | 1.000 | -0.0082 | 0.978 | 0.206 | 1.033 |
| Risk attitude in year 2003 (n=6077) ¹ | 22.757 | 20.034 | 22.471 | 19.796 | 29.952 | 24.238 |
| Human capital characteristics | | | | | | |
| Secondary degree with academic orientation (dummy) | 0.522 | 0.500 | 0.5235 | 0.499 | 0.4799 | 0.500 |
| Secondary education, GPA (scale 1-10) | 6.957 | 0.600 | 6.957 | .6005 | 6.958 | .5819 |
| # languages studied in secondary school | 2.864 | .825 | 2.8639 | .8237 | 2.886 | .8616 |
| # science subjects studied in secondary school | 2.295 | 1.3494 | 2.3024 | 1.3499 | 2.110 | 1.3236 |
| Tertiary degree Msc (dummy) | 0.513 | 0.500 | 0.5171 | 0.500 | 0.4154 | 0.493 |
| Tertiary education, GPA (scale 1-10) | 7.177 | .529 | 7.172 | .5268 | 7.275 | .583 |

¹Risk attitude is measured in terms of the stated reservation price to participate in a lottery where the bet is winning 1000 euro with a probability of 10%. Thus, risk aversion would imply a reservation price below 100. The question is included in the questionnaire of 2003 only, see Section 4.2.

4. Estimation results

In this section we will first test the propositions resulting from the model. We will then test the assumptions underlying the model and perform robustness checks.

¹⁶ The table suggests that the likelihood of being self-employed is higher for males who are older (in a range from 21 to 30), have lower levels of risk aversion, parents with higher education levels and a bachelor degree with high grades.

4.1 *Testing of propositions: The effect of the wage distribution in the labor market segment on the choice for self-employment*

We have estimated four specifications, determined by excluding outlying segments or not, combined with scaling the distribution moments or not. In Table 6, we estimate the choice equation at the aggregate level of field of education. The dependent variable, $P[SE]_k$, is the fraction in segment k that is self-employed 20 months after graduation. Control variables are average values per labor market segment. Table 7 shows the results of a probit model at the individual level, with standard errors clustered by labor market segments. The effects shown are marginal effects, i.e. the percentage increase in the probability of self-employment, $P[SE]_i$, when increasing the regressor by one unit (measured for an individual in the sample with average values for all regressors). As controls we use labor market dummies (regions and years) and personal characteristics (gender, age and parental education levels).^{17 18}

Our preferred specification excludes the four outlying segments, uses unscaled moments and is estimated on individual data using the controls specified in the previous section. Excluding the segments mitigates the multicollinearity problem, unscaled moments are closest to the specification we derived in Section 2 and individual data are most suited to control for individual characteristics. As Table 7, panel B, shows, our predictions are fully supported. The first and the third moment of the wage distribution have a significantly negative effect on the inclination to choose self-employment, the second moment has a significant positive effect. As the corresponding results in Table 7, panel A show, the results are not affected by including controls. If we estimate at the aggregate level of the labor market segment (Table 6) we get essentially the same results, whether we include controls or not (except for loss of significance in one case).

Scaling the distribution parameters does not affect the key conclusions: we still get significant confirmation of the predicted signs. The coefficients of the first moment decline by roughly 40%. The coefficients of the other moments increase, as should be expected: if the value of a regressor falls from dividing by a scaling factor, the value of the estimated coefficient will go up. Across all the specifications, the coefficient of the second moment increases roughly by a factor 40 in absolute value and the coefficient of the third moment by a factor 20. Judged by the mean values of the moments in Table 3, one would predict increases by a factor 25 and 15, respectively.¹⁹ Of course, the whole purpose of the rescaling is to get away from multicollinearity and not to rescale by the same constant for all observations. But the calculation suggests that the effect of rescaling is more or less in the ballpark that one might expect and that reducing multicollinearity does not dramatically upset parameter estimates.

The effect of excluding the four extreme labor market segments on estimated parameter values is very small and reduces the significance level in just a few cases, as can be seen from all pairwise comparisons.

¹⁷ Risk attitude is added as a regressor to the equation in a later stage when testing our assumptions. Its inclusion reduces the size of the sample to ten percent only due to the fact that it has been measured in year 2003 only. The set of human capital variables will be added to the equation later as well.

¹⁸ Due to the low number of observations at the aggregate level of labor market segments, we do not include the sets of average labor market characteristics and average personal controls simultaneously into the regression equations.

¹⁹ The mean value of σ_w^2 is 6, the value of σ_w / μ_w at mean values is 2.4/10; thus rescaling at these mean values would predict an increase by a factor 25.

This confirms the conclusion from the scaling exercise: we have no indications that multicollinearity renders our estimates essentially unreliable.

| Panel A No controls | | | | | | | | | |
|--|------------------|------------------|----------------|------------------|------------------|------------------|----------------|------------------|--|
| Dependent: $P/SE _k$ | k=118 | | | | k=114 | | | | |
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | | |
| | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | |
| μ_w | -.0336*** | .0067 | -.0182*** | .0058 | -.0297*** | .0064 | -.0179*** | .0055 | |
| σ_w^2 | .0231*** | .0035 | .8154*** | .1179 | .0182*** | .0038 | .7012*** | .1296 | |
| κ_w^3 | -.0010*** | .0002 | -.0213*** | .0056 | -.0008*** | .0003 | -.0167*** | .0053 | |
| Prob > F | 0.000 | | 0.000 | | 0.000 | | 0.0000 | | |
| R ² | 0.3292 | | 0.2976 | | 0.2163 | | 0.2415 | | |
| Panel B Labor market characteristics included as controls | | | | | | | | | |
| Dependent: $P/SE _k$ | k=118 | | | | k=114 | | | | |
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | | |
| | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | |
| μ_w | -.0361*** | .0070 | -.0206*** | .0063 | -.0315*** | .0067 | -.0195*** | .0059 | |
| σ_w^2 | .0224*** | .0034 | .8159*** | .1184 | .0178*** | .0037 | .6785*** | .1290 | |
| κ_w^3 | -.0010*** | .0002 | -.0191*** | .0059 | -.0007** | .0003 | -.0139** | .0055 | |
| Prob > F | 0.000 | | 0.000 | | 0.0000 | | 0.0000 | | |
| R ² | 0.4404 | | 0.4084 | | 0.3681 | | 0.3807 | | |
| Panel C Personal characteristics included as controls | | | | | | | | | |
| Dependent: $P/SE _k$ | k=118 | | | | k=114 | | | | |
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | | |
| | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | |
| μ_w | -.0383*** | .0080 | -.0219*** | .0066 | -.0323*** | .0077 | -.0221*** | .0065 | |
| σ_w^2 | .0171*** | .0035 | .6280*** | .1261 | .0126*** | .0037 | .4968*** | .1337 | |
| κ_w^3 | -.0006** | .0002 | -.0115** | .0055 | -.0004 | .0003 | -.0090* | .0052 | |
| Prob > F | 0.0000 | | 0.0000 | | 0.0000 | | 0.0000 | | |
| R ² | 0.4655 | | 0.4329 | | 0.3670 | | 0.3703 | | |

The p -values are based on robust standard errors. A significant coefficient at the 10% (5%) [1%] level is denoted by * (**) [***]. A constant term is included in all equations.

4.2 Testing of assumptions and robustness checks

Definition of labor market segments

As noted in the introduction, the validity of our approach hinges on a sensible definition of labor market segments. Table 3 already indicated that the wage variance within labor market segments is smaller than the variance between segments, as one would expect with sensibly defined labor market segments, see the Introduction. As a further check on the sensibility of the defined labor market segments, we repeated the analyses for 118 ‘labor market segments’ that are formed by random assignment of individuals to segments a

hundred times. For these ‘segments’ we didn’t establish any of the joint significant effects of the wage distribution in a segment on the occupational choice of individuals within that segment. The relationship between self-employment and the moments of the wage distribution is not some mechanical relationship that would hold in any arbitrary decomposition of the labor force.

| Panel A Dprobit, clustered by labor market segments, no controls | | | | | | | | | |
|--|------------------|------------------|----------------|------------------|-------------------|------------------|----------------|------------------|--|
| Dependent: <i>dummy Self-emp</i> | | | | | | | | | |
| $k=118, n=56,138$ | | | | | $k=114, n=53,819$ | | | | |
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | | |
| | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | |
| μ_w | -.0244*** | .0074 | -.0157*** | .0059 | -.0223*** | .0075 | -.0122*** | .0042 | |
| σ_w^2 | .0146*** | .0039 | .5345*** | .1335 | .0142*** | .0050 | .5747*** | .1768 | |
| κ_w^3 | -.0007*** | .0002 | -.0158*** | .0044 | -.0006*** | .0002 | -.0129*** | .0040 | |
| Prob > χ^2 | 0.0000 | | 0.0004 | | 0.0120 | | 0.0032 | | |
| Pseudo R ² | 0.0546 | | 0.0545 | | 0.0380 | | 0.0439 | | |
| Panel B Dprobit, clustered by labor market segments, labor market and personal characteristics included | | | | | | | | | |
| Dependent: <i>dummy Self-emp</i> | | | | | | | | | |
| $k=118$ | | | | | $K=114$ | | | | |
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | | |
| | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | |
| μ_w | -.0263*** | .0067 | -.0171*** | .005147 | -.0242*** | 0.0064 | -.0138*** | 0.0037 | |
| σ_w^2 | .0136*** | .0034 | .5274*** | .11179 | .0136*** | 0.0042 | .5590*** | 0.1527 | |
| κ_w^3 | -.0006*** | .0002 | -.0127*** | .00376 | -.0005*** | 0.0002 | -.0104*** | 0.0036 | |
| Prob > χ^2 | 0.0000 | | 0.000 | | 0.0000 | | 0.0000 | | |
| Pseudo R ² | 0.0739 | | 0.0754 | | 0.0593 | | 0.0635 | | |

The p -values are based on robust standard errors. A significant coefficient at the 10% (5%) [1%] level is denoted by * (**) [***]. A constant term is included in all equations.

The role and independence of self-employment income distributions in labor market segments

We have taken the approach of measuring the effect of wage income distributions only, instead of also including self-employment income distributions, because the literature has acknowledged that the latter are difficult to observe, for various reasons. These reasons apply in our case too. Only a small fraction of the sample is self-employed (3.6%), and a large fraction of the self-employed does not reveal their income at all. As a consequence, only 50 (19) out of the 118 labor market segments include at least 10 (20) observed incomes of self-employed. Moreover, the incomes show a large variance and this may reflect substantial measurement error. We have argued that individuals do not base their occupational choice on the self-employment income distribution within labor market segments, but on the self-employment income distribution in general.

Table 8 and Table 9 show how our results change upon including the first three central moments of p into the regression equations. The samples are limited to the 50 labor market segments in which there are at least 10 observations of p and to 47 when the sample of $k=114$ forms the basis. Table 8 shows the results at

the level of labor market segments k (comparable to Table 6), whereas Table 9 shows the results at the individual level n (comparable to Table 7). The bottom half of both tables shows the benchmark results, i.e., when only considering the labor market segments with at least ten observations of p but where the moments of p remain excluded.

| | $k=50$ | | | | $k=47$ | | | |
|--------------|------------------|------------------|----------------|------------------|------------------|------------------|----------------|------------------|
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | |
| | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> |
| μ_w | -.0476*** | .0144 | -.0238* | .01227 | -.0408*** | .0139 | -.0258** | .0113 |
| σ_w^2 | .0287*** | .0068 | .9253*** | .2307 | .0166* | .0098 | .8571*** | .2762 |
| κ_w^3 | -.0013*** | .0004 | -.0249 | .0175 | -.0004 | .0010 | -.0083 | .0168 |
| μ_p | -.0082 | .0054 | -.0094** | .0041 | -.0056 | .0052 | -.0046 | .0040 |
| σ_p^2 | 3.55e-06 | .0005 | .0094 | .0706 | -.0002 | .0005 | -.0504 | .0648 |
| κ_p^3 | 2.10e-06 | 6.15e-06 | .0161 | .0128 | 3.31e-06 | 6.61e-06 | .0210* | .0121 |
| Prob > F | 0.0000 | | 0.0000 | | 0.0109 | | 0.0019 | |
| R^2 | 0.4639 | | 0.4590 | | 0.3269 | | 0.3938 | |
| | $k=50$ | | | | $k=47$ | | | |
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | |
| | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> |
| μ_w | -0.0567*** | 0.0127 | -0.033** | 0.0116 | -0.0463*** | 0.0128 | -0.0305*** | 0.0106 |
| σ_w^2 | 0.0313*** | 0.0063 | 1.0617*** | 0.2403 | 0.0193** | 0.0092 | 0.8383*** | 0.2751 |
| κ_w^3 | -0.0014*** | 0.0004 | -0.0317* | 0.0175 | -0.0006 | 0.0010 | -0.0101 | 0.0167 |
| Prob > F | 0.0000 | | 0.0002 | | 0.0022 | | 0.0012 | |
| R^2 | 0.4270 | | 0.3426 | | 0.2854 | | 0.3059 | |

The tables show that our essential conclusion on the signs of the moments is not affected by the inclusion of the p -moments into the equations but that we lose precision, as one might expect with strong reduction in the number of observations. In addition, the moments of p themselves hardly affect the occupational choices of individuals. This lack of results is in line with previous findings and with the model we have proposed.

The results shown in Tables 8 and 9 include only those labor market segments with sufficient numbers of entrepreneurs, i.e., $n_p \geq 10$. This may lead to a biased sample excluding labor market segments with few entrepreneurs. In Table 10, we check our results for the full sample with self-employment income measures calculated for 13 segments. The segments are aggregates, for self-employment income only, of fields of education that are intrinsically related (e.g. fields in engineering, in teaching, in arts) and add up to sufficient number of self-employed to calculate moments of the income distribution (see the second column of tables A1 and A2). Thus, the underlying assumption is now that individuals assess self-

employment income from a set of related fields, instead of lumping all fields together (for instance, business economics, creative therapy, electrical engineering, etc).

Table 9

Adding the first three moments of the entrepreneurial income distribution per segment at level n

| Panel A | k= 50, n = 29,019 | | k=47, n = 27,398 | | k= 50, n = 29,019 | | k=47, n = 27,398 | |
|-----------------------|-------------------|-----------|------------------|-----------|-------------------|-----------|------------------|-----------|
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | |
| | Coeff. | Std. Err. | Coeff. | Std. Err. | Coeff. | Std. Err. | Coeff. | Std. Err. |
| μ_w | -.0318*** | .0096 | -.0168** | .0072 | -0.0269*** | 0.0096 | -0.0126** | 0.0063 |
| σ_w^2 | .0168*** | .0044 | .5476*** | .1466 | 0.0128* | 0.0078 | 0.7053*** | 0.1923 |
| κ_w^3 | -.0007*** | .0002 | -.0111 | .0098 | -0.0001 | 0.0006 | -0.0067 | 0.0083 |
| μ_p | -.0066* | .0038 | -.0071** | .0023 | -0.0051 | 0.0036 | -0.0004* | 0.0022 |
| σ_p^2 | -.00004 | .0003 | -.0112 | .0484 | -0.0001 | 0.0003 | -0.0547 | 0.0388 |
| κ_p^3 | 2.08e-06 | 3.08e-06 | .0145* | .0087 | 1.97E-06 | 3.25E-06 | 0.0213 | 0.0074 |
| Prob > χ^2 | 0.0000 | | 0.0000 | | 0.0013 | | 0.0000 | |
| Pseudo R ² | 0.0670 | | 0.0733 | | 0.0527 | | 0.0657 | |
| Panel B | k= 50, n = 29,019 | | k=47, n = 27,398 | | k= 50, n = 29,019 | | k=47, n = 27,398 | |
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | |
| | Coeff. | Std. Err. | Coeff. | Std. Err. | Coeff. | Std. Err. | Coeff. | Std. Err. |
| μ_w | -.0421*** | 0.0114 | -.0283*** | 0.0101 | -.0348*** | 0.0112 | .0209*** | 0.0069 |
| σ_w^2 | .0205*** | 0.0053 | .7433*** | 0.2276 | .0158** | 0.0080 | .7569*** | 0.2700 |
| κ_w^3 | -.0009*** | 0.0003 | -.0228* | 0.013 | -.0003* | 0.0006 | -.0087 | 0.0101 |
| Prob > χ^2 | 0.0000 | | 0.001 | | 0.0023 | | 0.0006 | |
| Pseudo R ² | 0.0612 | | 0.0556 | | 0.0474 | | 0.0506 | |

When we include the first three central moments of the entrepreneurial income distribution from the 13 segments, we get basically the same results as before. The moments of the wage distribution maintain their predicted signs when we include self-employment income moments, the magnitudes of the coefficients barely change and the self-employment income moments themselves have no significant effect. Only the first moment of the wage distribution loses significance, in case of scaled moments. Once again we find significant effects of opportunity cost and no effects of the self-employment income distribution.²⁰

²⁰ We also noted that our estimates are still unbiased if the distribution of self-employment income within a segment is independent of the distribution of wages. In all other cases, our estimates may be biased. The bias would be downwards in the case of a positive correlation and upwards in the case of a negative correlation. Table 11 shows our estimates, necessarily based on a modest number of observations. The correlations are positive, but quite small. Comparing the results in Tables 6 and 7 with those in 8, 9 and 10 indicates that including the p -moments in large majority leads to higher estimated coefficients of the moments of the wage distribution moments. This is in line with underestimation when we omit variables that correlate positively. When we compare the point estimates, in Table 8 and Table 9, for estimation excluding and including self-employment income, their ratio's are mostly around 1.2 to 1.3, which would be compatible with omitted variable bias at equal coefficients and the correlations as measured in Table 11. We conclude that the results we have presented for the full sample, if biased, are mostly likely an underestimate of the true effect of opportunity costs.

Table 10
Adding the moments of the p -distribution with an alternative definition of labor market segments

| Dprobit, clustered by alternative labor market segments, moments of p added to panel B of Table 7 | | | | | | | | |
|---|----------------------------------|------------------|----------------|------------------|----------------------------------|------------------|----------------|------------------|
| Dep: <i>dummyEntr</i> | $k=118, 13$ clusters, $n=56,138$ | | | | $k=114, 13$ clusters, $n=53,819$ | | | |
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | |
| | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> |
| μ_w | -.01934** | .0096 | -.0092 | .0062 | -.0176* | .0092 | -.0085 | 0.0063 |
| σ_w^2 | .0126*** | .0037 | .4607*** | .1088 | .0127*** | .0045 | .5403*** | 0.1666 |
| κ_w^3 | -.0005*** | .0002 | -.0120*** | .0043 | -.00050*** | .0002 | -.0116*** | 0.0040 |
| μ_{p_alt} | -.0046 | .0067 | -.0063 | .0051 | -.0039 | .0074 | -.0030 | .0062 |
| $\sigma_{p_alt}^2$ | -.00001 | .0011 | .0283 | .0495 | -.0000 | .0011 | .0132 | .0521 |
| $\kappa_{p_alt}^3$ | -4.98e-07 | .0001 | -.0004 | .0049 | 6.02e-07 | .0000 | -.0013 | .0043 |
| Prob > χ^2 | 0.0000 | | 0.0000 | | 0.0170 | | 0.0000 | |
| <i>Pseudo R</i> ² | 0.0566 | | 0.0612 | | 0.0400 | | 0.0456 | |

Table 11
Correlations between the moments of the w - and p -distributions in labor market segments*

| | $k=118$ labor market segments (50) | $k=114$ labor market segments (47) |
|---------------------|------------------------------------|------------------------------------|
| μ | 0.396 | 0.382 |
| σ^2 | 0.197 | 0.205 |
| κ^3 | 0.232 | 0.113 |
| σ/μ | 0.213 | 0.026 |
| κ^3/σ^3 | 0.018 | -0.120 |

*The correlations that involve an element of the distribution of entrepreneurial incomes have been calculated for the sample of 50 (47) labor market segments (out of the original 118 or 114) in which there are at least 10 entrepreneurial incomes observed.

4.3 Potential sources of bias

Biased moments

In our statistical analysis we use moments of the wage distribution as observed in the data. The distribution is realized after choices have been made and an individual may have private information on the distribution that will apply in his/her particular case. Correction for selectivity bias may be essential if one wants to retrieve the true risk that an individual is facing but the key question here is to what extent this is relevant for the individual's decision making. Econometric corrections for selectivity bias are often quite sensitive to specification and generally not very robust (eg. Chen, 2008; Mazza, Van Ophem and Hartog, 2010) but direct assessments of individuals' information on future incomes point in a single direction: there is no evidence that individuals systematically correct observations for private information (see for example Betts, 1996;

Dominitz and Manski, 1996; Brunello, Lucifora and Winter-Ebmer, 2004; Webbink and Hartog, 2004; Wolter and Weber, 2004; Schweri, Hartog and Wolter, 2011).

Omitted variables

Omitted variables, if correlated with a regressor, will lead to a biased estimate of the effect of that regressor. Variables that directly come to mind are risk attitude and human capital or ability. Both are known to affect the choice for entrepreneurship, are likely to affect the choice for a degree field and are also likely to affect (preferred) expected levels and variances of the wage distribution.

A comparison of Panel A with Panel B in Table 7 may be considered a first test of the possible effects of the omission of risk attitude. It shows that including controls for, among others, gender and age hardly affect the estimates of the coefficients of the first three moments of the wage distribution. Males are known to be less reluctant to take risks and risk aversion increases when people get older (Hartog et al., 2002). The estimates stand this first test.

A second test exploits the fact that our dataset includes an indicator of (stated) risk preference. Respondents were asked to value participation in a hypothetical lottery paying out €1000 with a 10 percent chance. Unfortunately, this indicator is available for only one out of the 10 waves of graduates in our sample (see Table 5 and its footnote). The average willingness to pay for participation in this hypothetical lottery reported by 6077 graduates was €22.8 (median €15), see Table 5. Thus, on average, the respondents in our sample are risk averse. The reported maximum was €99. The reservation price for participating in such a hypothetical lottery has been shown to be a valid (inverse) indicator of risk aversion and behavior under risk (see Cramer et al., 2002; Dohmen et al., 2011; Hartog et al., 2002). Table 5 shows that the average reservation price in the sample of entrepreneurs is significantly higher than in the sample of employees (€29.95 versus €22.47) in line with earlier applications (Cramer et al., 2002).

We replicate the analysis of Table 7 for this single wave of graduates and then include the stated risk preference variable as an additional regressor in the equation, see Panel A of Table 7 and Table 12.²¹ The results show a clear positive correlation between self-employment and the lottery reservation price. But what's more important, the estimated coefficients for the wage moments are not affected by the inclusion of a (strongly significant) risk preference variable; the magnitudes and standard errors only marginally different from the ones in Table 7. We conclude that our results are not likely to be driven by omitting a measure of risk attitude.²²

²¹ The replicated results for this wave compare well to the results presented in Table 7. Significance levels drop slightly due to the smaller sample, but conclusions remain unaltered. These results are not shown in Table 12.

²² Actually, our model indicates that risk attitude should not be included as a separate regressor: risk attitude is reflected in the regression coefficients for variance and skew. Thus, the key issue is not an omitted variable problem but coefficient heterogeneity. In the single wave we have data for, we interacted variance and risk attitude. Results were in the right direction (smaller effects for the less risk averse) but statistically weak.

Table 12

The effect of controlling for a measure of risk attitude at the individual level

| A risk attitude measure added to Panel B of Table 7 | | | | | | | | | |
|--|--------------------|------------------|----------------|------------------|--------------------|------------------|----------------|------------------|--|
| Dependent: <i>dummy Entr</i> | $k=118, n = 6,077$ | | | | $k=114, n = 5,835$ | | | | |
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | | |
| | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | |
| μ_w | -.0296*** | .0092 | -.0187*** | .0075 | -0.0283*** | 0.0093 | -0.0159*** | 0.0056 | |
| σ_w^2 | .0168*** | .0044 | .6018*** | .1594 | 0.0168*** | 0.0059 | 0.6314*** | 0.2216 | |
| κ_w^3 | -.0008*** | .0002 | -.0183*** | .0057 | -0.0007*** | 0.0003 | -0.0147*** | 0.0053 | |
| Lottery reservation price | .0006*** | .0001 | .0006*** | .0001 | 0.0006*** | 0.0001 | 0.0006*** | 0.0001 | |
| Prob > χ^2 | 0.0000 | | 0.000 | | 0.0000 | | 0.0000 | | |
| <i>Pseudo R</i> ² | 0.1087 | | 0.1034 | | 0.0842 | | 0.0880 | | |

The p -values are based on robust standard errors. A significant coefficient at the 10% (5%) [1%] level is denoted by * (**) [***]. A constant term is included in all equations.

Another stylized fact from the empirical entrepreneurship literature indicates that ability or, more in general human capital, affects the choice for entrepreneurship (Parker, 2009). Human capital might also induce people to choose certain degree fields and thereby affect the income distribution. Thus, human capital is the second obvious candidate to possibly confound our results. Table 13 shows that this is not the case. The analysis shown in Table 7 Panel B is repeated with inclusion of a set of human capital factors.²³ The estimated coefficients of the three measures characterizing the wage distribution show the same pattern as before.

Table 13

The effect of controlling for measures of human capital at the individual level

| Human capital controls added to panel B of Table 7 | | | | | | | | | |
|---|----------------------|------------------|----------------|------------------|--------------------|------------------|----------------|------------------|--|
| Dependent: <i>dummy Entr</i> | $k= 118, n = 56,138$ | | | | $k=114, n =53,819$ | | | | |
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | | |
| | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | |
| μ_w | -.0259*** | .0071 | -.0158*** | .0053 | -0.0241*** | 0.0067 | -0.0130*** | 0.0044 | |
| σ_w^2 | .0129*** | .0032 | .4983*** | .1031 | 0.0130*** | 0.0040 | 0.5290*** | 0.1454 | |
| κ_w^3 | -.0005*** | .0002 | -.0122*** | .0035 | -0.0004** | 0.0002 | -0.0102*** | 0.0036 | |
| <i>Pseudo R</i> ² | 0.0771 | | 0.0792 | | 0.0619 | | 0.066 | | |

The p -values are based on robust standard errors. A significant coefficient at the 10% (5%) [1%] level is denoted by * (**) [***]. A constant term is included in all equations. The human capital controls included here are listed in Table 5 (type of degree, grades, courses).

Field of education as a source of bias

Our data use income distribution moments grouped by type of education (academic discipline, field of education). Hence, we observe wage data for individuals who have chosen a particular field of education and

²³ The most obvious measures of human capital (education level and experience), are already controlled for in our sample of graduates from tertiary education at the start of their career.

we relate self-employment status to wage observations belonging to these preselected fields. What may go wrong here?

Consider taste for risk. Other things equal, a risk lover will choose an education track generating incomes with high variance more often. Risk lovers will also choose self-employment more often. Then self-employment will correlate positively with income variance across educations. We find indeed a positive correlation between self-employment and wage variance. However, the argument requires furthermore that the variance in the wage and self-employment income distributions correlate positively, for which we find only weak evidence in our data, see Table 11. Another observation required for the validity of this argument, to the extent that the labor market compensates for risk taking (Hartog, 2011), is that variance will be correlated positively with the mean income in a segment. This would generate a positive correlation between self-employment and mean income, contrary to what we find. As risk lovers would be attracted to high skew educations, we would also find positive correlation between skew and self-employment, again contrary to what we find. Thus, this potential spurious link runs counter to two of our three key findings.

Consider an omitted relevant ability that stimulates self-employment (e.g. the ability for fast and robust decision making). If this ability also stimulates to choose an education field with higher productivity as a wage earner, we would see a positive correlation between mean wage and self-employment; we find the opposite. The higher ability may reduce wage variance, as abler workers make fewer errors and have better foresight of pitfalls and dangers of failure; this would again predict the opposite of what we find, a negative correlation between self-employment and variance.

Omitted moments

In the introduction we noted that previous studies may have found no effect of income on occupational choice because they only include one (or two) moments of the relevant income distribution. We test this presumption for the specifications of Panel B of Table 7, see Table 14.

| Table 14 | | | | | | | | | |
|--|--------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|--|
| Including fewer moments of the wage distribution at the individual level | | | | | | | | | |
| Excluding κ_w^3 from the equation of panel B, Table 7 | | | | | | | | | |
| Dependent: <i>dummy Entr</i> | K= 118, n = 56,138 | | | | k=114, n =53,819 | | | | |
| | Unscaled moments | | Scaled moments | | Unscaled moments | | Scaled moments | | |
| | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | <i>Coeff.</i> | <i>Std. Err.</i> | |
| μ_w | -.0204*** | .0075 | -.0133*** | .0053 | -0.0202*** | 0.0063 | -0.0113*** | 0.0037 | |
| σ_w^2 | .0046*** | .0015 | .3898*** | .1025 | 0.0091*** | 0.0034 | 0.4339*** | 0.1545 | |
| <i>Pseudo R</i> ² | 0.0526 | | 0.0630 | | 0.0525 | | 0.0542 | | |
| Excluding σ_w^2 and κ_w^3 from the equation of panel B, Table 7 | | | | | | | | | |
| Dependent: <i>dummy Entr</i> | K= 118, n = 56,138 | | | | k=114, n =53,819 | | | | |
| | <i>Coeff.</i> | | <i>Std. Err.</i> | | <i>Coeff.</i> | | <i>Std. Err.</i> | | |
| μ_w | -0.0032 | | .0067 | | -0.0116** | | .0055 | | |
| <i>Pseudo R</i> ² | 0.0167 | | | | 0.0234 | | | | |

The *p*-values are based on robust standard errors. A significant coefficient at the 10% (5%) [1%] level is denoted by * (**) [***]. A constant term is included in all equations. The human capital controls included here are listed in Table 5.

The first moment has no effect on status choice if we omit the second and the third moment, just as in the earlier literature. As the top panel shows, substantial change in the estimated coefficient of the first moment only emerges if both the second and the third moment are omitted.

5. Conclusion

We set out to provide an understanding of the remarkable lack of support in the self-employment literature for the effect of financial returns on the decision to select self-employment above wage employment. Various explanations have been put forth for this puzzling lack of support. It is hard to believe that preferences for income entirely cease to play a role for the labor market choice between salaried and self-employment.

One explanation is that income data for the self-employed are poor and difficult to compare to wage earnings. Another explanation is that, besides the mean, variance and skew of the earnings distributions are also different and may drive the choice between these options. We try to address these empirical issues with a sample of recent graduates from tertiary education in the Netherlands. This sample is homogeneous with respect to age, education level and labor market experience and can be divided, based on degree fields, into labor market segments with their own wage distributions.

We find robust support for the effect of opportunity cost (moments of the wage distribution) and no significant effect of the financial returns to self-employment. Lower mean, higher variance and lower skew of the wage distribution for an individual's type of education increase the probability of self-employment. Robustness checks suggest that the conclusion is quite solid and that our estimates are more likely to give an underestimate than an overestimate. Lack of support for the moments of the self-employment income distribution is compatible with measurement error blurring estimated coefficients. It is also compatible with a blurred vision of the subjects themselves. The significant effects of opportunity cost of self-employment do indicate however, that the choice for self-employment is guided by financial incentives. Empirical research seeking to distinguish econometric problems of estimation under measurement errors and ignorance cum indifference among subjects would be a challenging next step.

References

- Astebro T., J. Mata and L. Santos-Pinto (2009), Preference for Skew in Lotteries: Evidence from the Laboratory?, *HEC Paris working paper*
- Astebro, T. (2010) 'The returns to entrepreneurship', chapter X in Cummings, D. (ed), *Handbook of Entrepreneurial Finance*, Oxford University Press
- Becker, G. (1964) *Human Capital: A theoretical and empirical analysis with special reference to education*, Chicago, University of Chicago Press (3rd edition, 1993)
- Benz, M. & B. Frey (2008), Being independent is a great thing: Subjective evaluations of self-employment and hierarchy, *Economica* 75(298), 362-383
- Betts, J.R.(1996).What do students know about wages? Evidence from a survey of undergraduates. *Journal of Human Resources*, 31(1), 27–56
- Bonin, H., T. Dohmen, A. Falk, D. Huffman and U. Sunde (2007), Cross-sectional earnings risk and occupational sorting: The role of risk attitudes, *Labour Economics* 14, 926-937
- Brunello, G.,Lucifora,C.,&Winter-Ebmer,R.(2004).The wage expectations of European college students. *Journal of Human Resources*, 39(4), 1116–1142
- Caliendo, M., F. Fossen and A. Kritikos (2009), Risk attitudes of nascent entrepreneurs—new evidence from an experimentally validated survey, *Small Business Economics* 32(2), 153-167
- Chen, S. (2008), Estimating the variance of wages in the presence of selection and unobserved heterogeneity, *Review of Economics and Statistics*, 90 (2), 275-289
- Cramer, J., J. Hartog, N. Jonker, and C.M. van Praag (2002), Low risk aversion encourages the choice for entrepreneurship: an empirical test of a truism, *Journal of Economic Behavior & Organization* 48(1), 29-36
- Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp and G. Wagner (2011), Individual risk attitudes: Measurement, determinants, and behavioral consequences, *Journal of the European Economic Association* 9(3), 522-550
- Dominitz, J. and C. Manski (1996), Eliciting student expectations of the return to schooling, *Journal of Human Resources*, 31, 1-26
- Dushnitsky, G. (2010), Entrepreneurial optimism in the market for technological invention, *Organization Science* 21(1), 150-167
- Garrett, T. and R. Sobel (1999), Gamblers favor skewness, not risk: further evidence from United States' Lottery games, *Economics Letters*, 63, 85-90.
- Golec, J. and M. Tamarkin (1998), Bettors love skewness, not risk, at the horse track, *Journal of Political Economy*, 106 (1), 205-225.
- Gollier, C. (2001), *The economics of risk and time*, Cambridge, Mass: The MIT Press
- Hamilton, B. (2000), Does entrepreneurship pay? An empirical analysis of the returns to self-employment, *Journal of Political Economy* 108(3), 604-631
- Hartog, J. (2011), A risk augmented Mincer earnings equation? Taking stock *Research in Labor Economics*, forthcoming
- Hartog, J., A. Ferrer Carbonell and N. Jonker (2002), Linking measured risk aversion to individual

- characteristics, *Kyklos*, 55(1), 3-26
- Hartog, J. and W. Vijverberg (2007), On compensation for risk aversion and skew affection in wages, *Labour Economics, Special Issue Education and Risk*, 14 (6), 938-956
- Hayward, M., D. Shepherd & D. Griffin (2006), A hubris theory of entrepreneurship, *Management Science* 52(2), 160-172.
- Hyytinen, A, P. Ilmakunnas and O. Toivanen (2008), The returns to entrepreneurship puzzle, Working paper
- Jacobs, B., J. Hartog and W. Vijverberg (2009) Self-selection bias in estimated wage premiums for earnings risk, *Empirical Economics* 37, 271–286
- Kawaguchi, D. (2003), Human capital accumulation of salaried and self-employed workers, *Labour Economics* 10, 55-71
- Kihlstrom, R. and J. Laffont (1979), A general equilibrium entrepreneurial theory of firm formation based on risk aversion, *Journal of Political Economy* 87, 719-749
- King, A. (1974), Occupational choice, risk aversion and wealth, *Industrial and Labor Relations Review*, 586-596
- Lazear, E. (1979) Why is there mandatory retirement?, *Journal of Political Economy* 87, 1261-1284
- Lowe, R. and A. Ziedonis (2006), Overoptimism and the performance of entrepreneurial firms, *Management Science* 52(2), 173-186
- Lucas, R. (1978), On the size distribution of business firms, *Bell Journal of Economics* 9, 508-23
- Macdonald, G. (1988), The economics of rising stars, *American Economic Review* 78(1), 155-166.
- Marshall, A. (1930), *Principles of Economics*, Macmillan and Co., London (first edition 1890).
- Mazza, J., H. van Ophem and J. Hartog (2010), Unobserved heterogeneity and risk in wage variance: does schooling provide earnings insurance?, Bonn: IZA DP 5531
- Nicolaou, N., S. Shane, L. Cherkas, J. Hunkin & T.D. Spector (2008), Is the tendency to engage in entrepreneurship genetic?, *Management Science* 54, 167-179
- Parker, S.C. (2009), *The Economics of Entrepreneurship*, Cambridge, UK: Cambridge University Press
- Parker, S and CM van Praag (2010) Group status and entrepreneurship, *Journal of Economics and Management Strategy* 19(4), 919–945
- Reich, M, D. Gordon and R. Edwards (1973) A Theory of Labor Market Segmentation, *The American Economic Review Papers and Proceedings* 63(2), 359-365
- Rosen, S. (1981) The Economics of Superstars. *American Economic Review* 71(5), 845-858
- Schweri, J., J. Hartog and S. Wolter (2011), Do students expect compensation for wage risk? , *Economics of Education Review*, 30 (2), 215-227
- Tsiang, S.C. (1972), The rationale for mean-standard deviation analysis, skewness preference and the demand for money, *American Economic Review*, 62 (3), 354-371.
- Webbink, D. and J. Hartog (2004), Can students predict their starting salary? Yes!, *Economics of Education Review*, 23 (2), 103-113
- Wolter, S. and B. Weber (2004), Returns to education: are students' expectations rational? Working Paper Swiss Coordination Center for Research in Education, Aarau, Switzerland

Table A1. Descriptive statistics of fields of study (Bachelor)

| Labor market segment | Grouped segm | N obs | Self | wage | mean | mean | N obs | Mean | N obs | |
|--------------------------------------|-----------------|-------|----------|----------|-----------|----------|-------|--------------|-------|--------------|
| | | | employed | variance | log(wage) | wage (€) | | Entr. income | | Entr. income |
| Business Economics/Business Sciences | Economics | 653 | 0.012 | 0.0262 | 0.0029 | 2.221 | 10.23 | 625 | 11.61 | 7 |
| Commerce | Economics | 586 | 0.031 | 0.0301 | 0.0050 | 2.215 | 9.60 | 554 | 11.37 | 15 |
| Business Informatics | Economics | 671 | 0.031 | 0.0278 | 0.0032 | 2.250 | 10.14 | 642 | 11.86 | 18 |
| Communication | Economics | 597 | 0.050 | 0.0314 | 0.0032 | 2.218 | 9.66 | 557 | 10.24 | 26 |
| Accountancy | Economics | 558 | 0.005 | 0.0315 | 0.0026 | 2.224 | 10.23 | 525 | 6.81 | 1 |
| International Business and Languages | Economics | 433 | 0.007 | 0.0275 | 0.0043 | 2.192 | 9.34 | 418 | 9.79 | 2 |
| Tourism & Leisure | Economics | 475 | 0.013 | 0.0338 | 0.0049 | 2.110 | 9.32 | 455 | 8.92 | 4 |
| Hotel Management | Economics | 531 | 0.026 | 0.0295 | 0.0036 | 2.190 | 9.59 | 509 | 9.50 | 13 |
| Small Business and Retail Management | Economics | 335 | 0.093 | 0.0356 | 0.0045 | 2.224 | 10.32 | 289 | 9.21 | 23 |
| Management, Economics & Law | Economics | 525 | 0.015 | 0.0280 | 0.0020 | 2.212 | 9.99 | 503 | 10.56 | 4 |
| Logistics & Economics | Economics | 649 | 0.014 | 0.0277 | 0.0035 | 2.217 | 10.17 | 620 | 8.81 | 7 |
| Facility Services | Economics | 612 | 0.023 | 0.0279 | 0.0034 | 2.191 | 9.63 | 596 | 14.19 | 13 |
| Journalism | Journalism | 587 | 0.177 | 0.0346 | 0.0010 | 2.248 | 10.24 | 508 | 11.15 | 85 |
| Business Management | Economics | 163 | 0.049 | 0.0209 | 0.0012 | 2.233 | 9.60 | 153 | 14.21 | 5 |
| Fiscal Economics | Economics | 269 | 0.019 | 0.0356 | 0.0025 | 2.258 | 10.27 | 260 | 7.42 | 4 |
| European professions | Economics | 177 | 0.028 | 0.0349 | 0.0042 | 2.209 | 9.63 | 164 | 8.88 | 4 |
| Leisure Management | Economics | 213 | 0.042 | 0.0352 | 0.0055 | 2.143 | 9.14 | 196 | 8.44 | 8 |
| International Business & Management | Economics | 73 | 0.055 | 0.0369 | 0.0021 | 2.223 | 9.79 | 66 | 4.48 | 2 |
| Real Estate | Economics | 121 | 0.033 | 0.0382 | 0.0041 | 2.191 | 9.18 | 119 | 3.36 | 1 |
| Marketing management | Economics | 69 | 0.058 | 0.0293 | 0.0033 | 2.205 | 9.44 | 64 | 10.71 | 3 |
| Personnel & Labour | Social sciences | 575 | 0.023 | 0.0284 | 0.0027 | 2.232 | 9.92 | 550 | 16.10 | 12 |
| Socio-Cultural Studies | Social sciences | 499 | 0.048 | 0.0342 | 0.0025 | 2.211 | 9.60 | 457 | 8.91 | 18 |
| Social Work & Services | Social sciences | 624 | 0.011 | 0.0218 | 0.0006 | 2.278 | 10.22 | 599 | 7.49 | 6 |
| Social Pedagogy | Social sciences | 908 | 0.013 | 0.0263 | 0.0021 | 2.217 | 9.78 | 858 | 8.88 | 11 |
| Socio-Legal Services | Social sciences | 414 | 0.010 | 0.0234 | 0.0027 | 2.255 | 10.10 | 397 | 8.39 | 4 |
| Information Management | Social sciences | 385 | 0.016 | 0.0311 | 0.0037 | 2.205 | 9.65 | 366 | 8.30 | 3 |
| Creative Therapy | Social sciences | 162 | 0.056 | 0.0369 | 0.0029 | 2.251 | 9.98 | 147 | 9.48 | 7 |
| Medical Laboratory Technician | (Para) medical | 569 | 0.005 | 0.0232 | 0.0042 | 2.157 | 9.04 | 547 | 8.10 | 3 |
| Nursing | (Para) medical | 882 | 0.010 | 0.0200 | 0.0018 | 2.241 | 10.07 | 847 | 9.76 | 8 |
| Physiotherapy | (Para) medical | 724 | 0.057 | 0.0284 | -0.0003 | 2.358 | 14.34 | 642 | 16.33 | 30 |
| Speech Therapy | (Para) medical | 568 | 0.067 | 0.0315 | 0.0033 | 2.239 | 11.42 | 489 | 18.40 | 32 |
| Nutrition & Dietetics | (Para) medical | 568 | 0.019 | 0.0295 | 0.0036 | 2.237 | 10.11 | 529 | 11.16 | 9 |

Table A1 (Continued) Descriptive statistics of fields of study (Bachelor)

| | | | | | | | | | | |
|--|----------------|-----|-------|--------|---------|-------|-------|-----|-------|-----|
| Ergotherapy | (Para) medical | 709 | 0.082 | 0.0258 | 0.0017 | 2.283 | 10.79 | 623 | 13.18 | 43 |
| Medical Imaging & Radiotherapy | (Para) medical | 259 | 0.004 | 0.0121 | 0.0012 | 2.212 | 9.19 | 252 | 8.23 | 1 |
| Oral Hygiene | (Para) medical | 105 | 0.057 | 0.0154 | 0.0004 | 2.339 | 11.79 | 94 | 26.40 | 3 |
| Environmental Management/Technology | Agriculture | 438 | 0.018 | 0.0287 | 0.0032 | 2.236 | 9.79 | 414 | 12.96 | 6 |
| Agri-Business | Agriculture | 365 | 0.033 | 0.0255 | 0.0031 | 2.237 | 9.64 | 334 | 11.51 | 11 |
| Animal Husbandry | Agriculture | 433 | 0.095 | 0.0480 | 0.0082 | 2.151 | 9.46 | 380 | 7.14 | 21 |
| Food Technology | Agriculture | 98 | 0.000 | 0.0284 | 0.0034 | 2.245 | 9.80 | 96 | | 0 |
| Horticulture & agriculture | Agriculture | 59 | 0.186 | 0.0392 | -0.0023 | 2.161 | 8.86 | 48 | 10.49 | 5 |
| Primary School Teacher | Teaching | 896 | 0.004 | 0.0190 | 0.0016 | 2.285 | 10.78 | 812 | 8.20 | 3 |
| Physical Education Teacher, Grade 1 | Teaching | 472 | 0.023 | 0.0364 | 0.0010 | 2.314 | 11.41 | 432 | 11.96 | 10 |
| Dutch Teacher | Teaching | 320 | 0.031 | 0.0320 | 0.0043 | 2.283 | 10.70 | 302 | 10.42 | 7 |
| Economics Teacher (general & business) | Teaching | 370 | 0.019 | 0.0271 | 0.0024 | 2.259 | 10.66 | 351 | 8.37 | 6 |
| Special Needs Teacher | Teaching | 392 | 0.000 | 0.0183 | 0.0013 | 2.299 | 10.33 | 356 | | 0 |
| Social Studies Teacher | Teaching | 93 | 0.022 | 0.0311 | 0.0064 | 2.200 | 9.24 | 91 | 13.11 | 2 |
| Education | Teaching | 276 | 0.011 | 0.0263 | 0.0012 | 2.256 | 10.53 | 256 | 29.50 | 3 |
| Science Teacher | Teaching | 452 | 0.007 | 0.0275 | 0.0017 | 2.316 | 10.92 | 434 | 6.37 | 2 |
| Geography/History Teacher | Teaching | 497 | 0.026 | 0.0419 | 0.0036 | 2.280 | 10.98 | 471 | 10.13 | 10 |
| Arts & Crafts Teacher | Teaching | 86 | 0.128 | 0.0540 | 0.0094 | 2.186 | 9.60 | 69 | 9.20 | 10 |
| English/French/German Teacher | Teaching | 603 | 0.013 | 0.0378 | 0.0030 | 2.294 | 11.58 | 558 | 8.59 | 6 |
| Visual Arts & Design | Teaching | 678 | 0.354 | 0.0453 | 0.0027 | 2.148 | 9.81 | 426 | 7.66 | 152 |
| Music | Arts | 329 | 0.368 | 0.0492 | -0.0001 | 2.288 | 12.42 | 179 | 11.71 | 68 |
| Theater | Arts | 98 | 0.204 | 0.0314 | 0.0027 | 2.130 | 9.28 | 72 | 11.76 | 12 |
| Dance | Arts | 60 | 0.200 | 0.1231 | 0.0050 | 1.976 | 9.13 | 39 | 9.50 | 9 |
| Chemical Technician | Engineering | 240 | 0.008 | 0.0306 | 0.0047 | 2.180 | 9.05 | 226 | 7.50 | 2 |
| Structural Engineering | Engineering | 506 | 0.020 | 0.0251 | 0.0041 | 2.206 | 9.35 | 474 | 8.98 | 9 |
| Electrical Engineering | Engineering | 453 | 0.026 | 0.0258 | 0.0039 | 2.247 | 9.75 | 430 | 8.36 | 8 |
| Civil Engineering | Engineering | 430 | 0.007 | 0.0271 | 0.0048 | 2.216 | 9.84 | 410 | 7.27 | 2 |
| Chemical Engineering | Engineering | 651 | 0.011 | 0.0236 | 0.0031 | 2.248 | 9.99 | 625 | 9.38 | 6 |
| Applied Informatics | Engineering | 862 | 0.057 | 0.0272 | 0.0031 | 2.250 | 10.11 | 788 | 10.55 | 38 |
| Mechanical Engineering | Engineering | 442 | 0.020 | 0.0306 | 0.0044 | 2.225 | 9.76 | 418 | 10.70 | 9 |
| Naval Officer | Engineering | 98 | 0.010 | 0.0756 | 0.0105 | 2.097 | 9.11 | 87 | 8.45 | 1 |
| Physics Engineering | Engineering | 99 | 0.010 | 0.0194 | 0.0019 | 2.243 | 9.59 | 91 | 16.80 | 1 |
| Fashion Management and Technology | Engineering | 160 | 0.081 | 0.0322 | 0.0049 | 2.160 | 10.17 | 145 | 9.35 | 9 |
| Car Technology | Engineering | 112 | 0.018 | 0.0242 | 0.0038 | 2.243 | 10.70 | 107 | 1.82 | 2 |

Table A2. Descriptive statistics of fields of study (Master)

| | Grouped segm | N obs | self employed | wage variance | wage skewness | mean log(wage) | mean | N obs | Mean | N obs |
|------------------------------|----------------------|-------|---------------|---------------|---------------|----------------|----------|-------|--------------|--------------|
| | | | | | | | wage (€) | wage | Entr. income | Entr. income |
| Dutch | Humanities/Languages | 549 | 0.0729 | 0.0267 | 0.0019 | 2.294 | 10.67 | 508 | 11.20 | 30 |
| English | Humanities/Languages | 441 | 0.0680 | 0.0362 | 0.0021 | 2.263 | 10.77 | 399 | 11.84 | 24 |
| Other languages | Humanities/Languages | 356 | 0.0534 | 0.0342 | 0.0018 | 2.252 | 10.81 | 328 | 9.08 | 15 |
| Philosophy/Theology | Humanities/Languages | 189 | 0.0899 | 0.0387 | 0.0000 | 2.299 | 10.52 | 162 | 10.73 | 12 |
| History | Humanities/Languages | 571 | 0.0490 | 0.0338 | 0.0007 | 2.277 | 10.59 | 527 | 11.11 | 23 |
| Language & Culture (general) | Humanities/Languages | 395 | 0.0582 | 0.0315 | 0.0021 | 2.255 | 10.09 | 374 | 9.79 | 19 |
| Art History & Archaeology | Humanities/Languages | 265 | 0.1019 | 0.0362 | 0.0035 | 2.221 | 9.44 | 235 | 8.67 | 19 |
| Corporate Communications | Humanities/Languages | 334 | 0.0120 | 0.0233 | 0.0015 | 2.255 | 10.01 | 323 | 15.02 | 4 |
| European Studies | Humanities/Languages | 68 | 0.0294 | 0.0225 | 0.0040 | 2.243 | 10.12 | 67 | 5.56 | 2 |
| Film, Television & Theatre | Humanities/Languages | 146 | 0.1301 | 0.0459 | 0.0041 | 2.235 | 10.59 | 123 | 11.12 | 14 |
| Alpha Information science | Humanities/Languages | 305 | 0.0492 | 0.0255 | 0.0013 | 2.286 | 10.46 | 286 | 9.21 | 9 |
| Chemistry | Techn/Engineering | 516 | 0.0039 | 0.0367 | 0.0043 | 2.199 | 9.64 | 502 | 7.92 | 2 |
| Computer Science | Techn/Engineering | 319 | 0.0439 | 0.0306 | 0.0022 | 2.301 | 10.42 | 298 | 14.51 | 10 |
| Biology | Techn/Engineering | 779 | 0.0282 | 0.0361 | 0.0015 | 2.193 | 9.62 | 748 | 11.12 | 16 |
| Pharmacy | Techn/Engineering | 503 | 0.0298 | 0.0244 | -0.0012 | 2.427 | 12.58 | 465 | 14.88 | 13 |
| Pure Mathematics/Physics | Techn/Engineering | 533 | 0.0131 | 0.0366 | 0.0027 | 2.224 | 9.98 | 519 | 9.05 | 6 |
| Agricultural Science | Techn/Engineering | 67 | 0.0299 | 0.0214 | 0.0021 | 2.337 | 10.86 | 63 | 9.00 | 1 |
| Chemi/Techn Agri-sciences | Techn/Engineering | 357 | 0.0280 | 0.0271 | 0.0019 | 2.277 | 10.19 | 332 | 8.73 | 7 |
| Bioprocessing & Food Tech | Techn/Engineering | 692 | 0.0159 | 0.0333 | 0.0016 | 2.259 | 10.23 | 660 | 9.21 | 9 |
| Architecture | Techn/Engineering | 820 | 0.0512 | 0.0213 | 0.0022 | 2.296 | 10.74 | 771 | 12.41 | 32 |
| Mechanical Engineering | Techn/Engineering | 747 | 0.0134 | 0.0217 | 0.0010 | 2.352 | 11.32 | 702 | 11.27 | 9 |
| Electrical Engineering | Techn/Engineering | 458 | 0.0240 | 0.0219 | 0.0008 | 2.345 | 11.30 | 428 | 17.08 | 8 |
| Chemical Engineering | Techn/Engineering | 555 | 0.0126 | 0.0291 | 0.0014 | 2.330 | 10.96 | 540 | 20.59 | 4 |
| Civil Engineering | Techn/Engineering | 720 | 0.0181 | 0.0180 | 0.0016 | 2.329 | 10.83 | 695 | 14.80 | 9 |
| Technology & Management | Techn/Engineering | 747 | 0.0281 | 0.0201 | 0.0005 | 2.385 | 11.59 | 696 | 9.03 | 19 |
| Industrial Design | Techn/Engineering | 416 | 0.1226 | 0.0240 | 0.0016 | 2.299 | 10.32 | 365 | 11.04 | 38 |
| Aerospace Engineering | Techn/Engineering | 146 | 0.0411 | 0.0218 | 0.0012 | 2.359 | 10.79 | 138 | 12.23 | 4 |
| Applied Computer Science | Techn/Engineering | 364 | 0.0632 | 0.0193 | 0.0012 | 2.328 | 10.73 | 331 | 9.85 | 19 |
| Applied Math/Physics | Techn/Engineering | 674 | 0.0089 | 0.0312 | 0.0013 | 2.287 | 10.40 | 647 | 13.06 | 5 |
| Economics | Economics&law | 1502 | 0.0220 | 0.0250 | 0.0007 | 2.349 | 11.32 | 1435 | 14.12 | 29 |
| Business Science | Economics&law | 808 | 0.0186 | 0.0241 | 0.0007 | 2.352 | 11.39 | 775 | 11.71 | 11 |
| Econometrics | Economics&law | 509 | 0.0138 | 0.0260 | -0.0006 | 2.360 | 11.75 | 492 | 11.90 | 6 |

Table A2 (Continued) Descriptive statistics of fields of study (Master)

| | | | | | | | | | | |
|-------------------------|-----------------|------|--------|--------|---------|-------|-------|------|-------|----|
| Fiscal Economics | Economics&law | 218 | 0.0138 | 0.0161 | 0.0001 | 2.420 | 11.72 | 208 | 9.82 | 1 |
| Business Administration | Economics&law | 835 | 0.0335 | 0.0224 | 0.0001 | 2.355 | 11.52 | 783 | 10.81 | 26 |
| Dutch Law | Economics&law | 1078 | 0.0176 | 0.0210 | 0.0010 | 2.341 | 11.29 | 1028 | 10.11 | 14 |
| Notarial Law | Economics&law | 467 | 0.0064 | 0.0273 | 0.0012 | 2.339 | 10.97 | 449 | 4.97 | 1 |
| Fiscal Law | Economics&law | 509 | 0.0118 | 0.0175 | 0.0000 | 2.414 | 12.38 | 494 | 20.89 | 4 |
| International Law | Economics&law | 71 | 0.0000 | 0.0275 | -0.0005 | 2.332 | 10.71 | 69 | | 0 |
| Healthcare | Health | 782 | 0.0128 | 0.0300 | 0.0002 | 2.316 | 10.90 | 740 | 13.26 | 8 |
| Medicine | Health | 998 | 0.0070 | 0.0237 | -0.0003 | 2.416 | 11.92 | 975 | 27.02 | 6 |
| Biomedical Sciences | Health | 610 | 0.0082 | 0.0274 | 0.0013 | 2.238 | 9.76 | 590 | 11.83 | 5 |
| Veterinary Science | Health | 278 | 0.0612 | 0.0232 | 0.0000 | 2.369 | 11.19 | 253 | 9.76 | 12 |
| Sociology | Social sciences | 491 | 0.0305 | 0.0294 | -0.0001 | 2.304 | 10.50 | 470 | 11.97 | 9 |
| Psychology | Social sciences | 1093 | 0.0357 | 0.0358 | 0.0001 | 2.305 | 10.92 | 1037 | 17.52 | 34 |
| Politics | Social sciences | 463 | 0.0454 | 0.0265 | -0.0004 | 2.341 | 11.21 | 440 | 10.78 | 18 |
| Education Science | Social sciences | 698 | 0.0143 | 0.0295 | 0.0004 | 2.336 | 11.73 | 661 | 13.45 | 8 |
| (Applied) Education | Social sciences | 380 | 0.0342 | 0.0232 | 0.0010 | 2.336 | 11.17 | 354 | 12.41 | 12 |
| Cultural Anthropology | Social sciences | 367 | 0.0436 | 0.0340 | 0.0018 | 2.249 | 10.64 | 347 | 12.48 | 15 |
| Communication | Social sciences | 705 | 0.0369 | 0.0260 | 0.0018 | 2.300 | 10.63 | 663 | 12.45 | 20 |
| Socio-Cultural Science | Social sciences | 779 | 0.0257 | 0.0259 | 0.0013 | 2.315 | 10.69 | 750 | 9.93 | 17 |
| Public Administration | Social sciences | 1080 | 0.0194 | 0.0223 | 0.0008 | 2.361 | 11.30 | 1032 | 12.13 | 19 |
| Human Geography & Plann | Social sciences | 1069 | 0.0178 | 0.0241 | 0.0008 | 2.309 | 10.47 | 1026 | 12.04 | 16 |