Labour Economics 20 (2013) 57–67

Contents lists available at SciVerse ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

The return-to-entrepreneurship puzzle☆

Ari Hyytinen [a,1], Pekka Ilmakunnas [b,*], Otto Toivanen [c,2]

[a] Jyväskylä University School of Business and Economics and Yrjö Jahnsson Foundation, Finland
[b] Aalto University School of Business, Finland
[c] Katholieke Universiteit Leuven, Belgium

HIGHLIGHTS

► The returns to entrepreneurship are monetary and non-monetary.
► We offer new evidence on these returns using a large sample of male twins.
► Entrepreneurs earn a negative wage premium.
► Panel estimates of the premium are upwards and OLS estimate downwards biased.
► Entrepreneurs have longer work hours and more responsibilities. However, entrepreneurs have a greater control over their work.

ARTICLE INFO

Article history:
Received 15 February 2012
Received in revised form 2 October 2012
Accepted 2 November 2012
Available online 9 November 2012

JEL classification:
L26
J24

Keywords:
Entrepreneurship
Earnings
Twin data
Monetary returns
Non-monetary returns
Selection

ABSTRACT

The returns to entrepreneurship are monetary and non-monetary. We offer new evidence on these returns using a large sample of male twins. Our within-twin analysis suggests that OLS estimates are downwards, and panel data estimates upwards biased. The within-twin estimates imply that entrepreneurs earn a negative earnings premium. Our within-twin analysis of non-monetary returns shows that entrepreneurs work longer hours and have greater responsibilities, but also have a greater control over their work.

1. Introduction

Hardly anyone thinks that the returns to entrepreneurship would not be both monetary and non-monetary. There is less agreement on how large those returns are. Some argue that there is a return-to-entrepreneurship puzzle, i.e., that despite working longer hours and bearing greater risks, entrepreneurs earn on average less than employees (see, e.g., Evans and Leighton, 1989; Carrington et al., 1996; Hamilton, 2000).[3] Not all analyses support this view (see, e.g., Fairlie, 1995; Berglann et al., 2011; Tergiman, 2011), and some others emphasize the non-monetary returns, such as greater independence and higher job

☆ We would like to thank Jaakko Kaprio, Edvard Lazear, Mirjam van Praag and seminar participants at HECER (Helsinki, 2008), EARIE (Toulouse, 2008), the 3rd Nordic Summer Institute in Labour Economics (Aarhus, 2008) and the XXX Annual Meeting of the Finnish Society for Economic Research (Jyväskylä, 2008), as well as Ian Walker (editor) and two referees for useful comments, Jaakko Kaprio for access to the twin data and Anni Nevalainen and Mari Rantanen for excellent research assistance. This research has been financially supported by the Academy of Finland (project 127796) and Toivanen's research has also been supported by grant GOA/12/003. Toivanen also thanks the Yrjö Jahnsson Foundation for support and Hyytinen the Bank of Finland for hospitality. The usual caveat applies.

* Corresponding author at: Department of Economics, Aalto University School of Business, P.O. Box 21240, FI-00076 Aalto, Helsinki, Finland.
E-mail addresses: ari.hyytinen@jyu.fi (A. Hyytinen), pekka.ilmakunnas@aalto.fi (P. Ilmakunnas), otto.toivanen@econ.kuleuven.be (O. Toivanen).
[1] Jyväskylä University School of Business and Economics, P.O. Box 35, FI-40014 University of Jyväskylä, Finland.
[2] Department of Managerial Economics, Strategy and Innovation, Katholieke Universiteit Leuven, Naamsestraat 69 Leuven, B-3000 Belgium.

[3] Evidence for the claim that the self-employed bear greater risk and face a more variable income stream than regular employees can be found from, e.g., Carrington et al. (1996).

0927-5371/$ – see front matter
http://dx.doi.org/10.1016/j.labeco.2012.11.002

satisfaction.[4] We provide new evidence on both monetary and non-monetary returns using a large sample of male twins. We also provide evidence that entrepreneurial dynamics may invalidate the use of traditional panel data estimators.

While the prior analyses are insightful, they are on balance inconclusive about the importance of the monetary and non-monetary returns to entrepreneurship. Besides differences in the sources of data and subtle measurement issues (see, e.g., Parker, 2009; Åstebro, 2011), there are two potential reasons for this somewhat unfortunate state of affairs: First, unobserved heterogeneity may hamper the inference. A number of the earlier findings are based on cross-sectional analyses that cannot tell apart the implications of unobserved heterogeneity, such as differences in ability (e.g. productivity at work; see also Åstebro et al., 2011), risk aversion, and willingness to substitute work for leisure. Second, even though there are studies that use panel data to control for the unobserved heterogeneity, they implicitly introduce another potential bias. This bias arises because the identification is based on within-individual (temporal) variation in the data.

Using the within-individual variation for the identification of the returns to entrepreneurship is problematic because of the dynamics of entrepreneurial entry and exit: Those who have just entered may be struggling for survival and thus working longer hours, earning less, and bearing greater risks than the (incumbent) entrepreneurs. Conversely, it could be that the early years of an entrepreneurial spell are the most rewarding (see Tergiman, 2011). New entrepreneurs may also enter entrepreneurship from a state that is unusually bad, for example from unemployment or from an unsatisfying job. The direction of the bias would then depend on the relative strength of these effects. Using entrepreneurial exits for identification is at least as problematic, because most of the entrepreneurial (self-employment) spells are very brief both in the US (see, e.g., Evans and Leighton, 1989; Bruce and Schuetze, 2004) and in Europe (see, e.g., Taylor, 1999; Hyytinen and Rouvinen, 2008). They are brief because many of the new entrepreneurs fail soon after entry. This means that in studies making use of panel data, within-individual variation at the end of an entrepreneurial spell may arise from the labor market switches of those who lost the battle for survival (whereas the most successful entrepreneurial spells are right-censored). It is also possible that exit from entrepreneurship may lead into a worse than average employment state. Standard panel estimators, such as the first-differencing (FD) or the fixed effects (FE) -estimators, identify returns to entrepreneurship from this type of variation or assume that contemporaneous or expected future shocks do not affect occupational choices.[5] We show that these worries are with some foundation in the data we use.

This paper offers new evidence on the returns to entrepreneurship using data from a large sample of genetically identical (monozygotic, MZ) twins that has been matched to linked employee–employer data.[6] These data allow us to focus on *within-twin pair* variation. Twin differencing leads to a within-twin (WT) estimator that controls for genetic factors that, we argue, are one of the prime sources of unobserved heterogeneity.[7] Supporting this view, Nicolaou et al. (2008a), Nicolaou and Shane (2009) and Shane et al. (2010) explain that the tendency of individuals to become entrepreneurs depends on genetics for four main reasons (see also Van der Loos et al., 2011; Van der Loos et al., 2010): First, certain genetic compositions have a direct effect on the chemical mechanisms in an individual's brain that may expose her to entrepreneurship (by, e.g., influencing how she perceives risk taking). Second, they contribute to the development of personality traits (e.g., extraversion) that almost certainly are beneficial in entrepreneurship (see also Nicolaou et al., 2008b). Third, certain genetic compositions make an individual more responsive to environmental stimuli (e.g. business opportunities). Finally, genetic factors affect the likelihood that an individual self-selects into environments supportive to entrepreneurship (leading to so called gene-environment correlation). Yet another reason why comparing (monozygotic) twins controls for the unobserved heterogeneity is that besides the same genetic environment, they share the same family background and typically experience more similar environments than for example non-twin siblings or children of different families do. This means that twin differencing is robust to, e.g., intergenerational correlation in self-employment (see, Dunn and Holtz-Eakin, 2000), as it does not rely on sample variation *between* families.[8]

We apply the OLS, FD, FE and WT estimators to a large panel data on Finnish male twins. Our long observation period makes it possible to observe relatively long entrepreneurial spells. Our FE estimator therefore allows us to estimate the average returns to entrepreneurship over such spells. The WT estimator achieves this same goal even with cross-section data because it utilizes data on entrepreneurs at different phases of their entrepreneurial spells. This is important because returns to entrepreneurship need not be the same in each year of entrepreneurship.

We find that the estimates of monetary returns can be ranked in a particular way, i.e., that $\hat{\gamma}_{OLS} < \hat{\gamma}_{WT} < \hat{\gamma}_{FE} < \hat{\gamma}_{FD}$. This ranking of the estimators is robust to a number of sensitivity checks and consistent with the interpretation that there is unobserved heterogeneity (pushing the OLS estimates downwards) and that using within-individual variation to control for it is hampered by the dynamics of entrepreneurial entry and exit. The data suggests, in particular, that the FD and FE estimators may be biased upwards, because they do not adjust for the possibility that there is a dip in the employment income before a switch to entrepreneurship (not unlike in Ashenfelter, 1978) and, again, after the entrepreneurial spell. The WT estimator, on the other hand, is more robust to unobserved heterogeneity and to income dips and suggests a negative earnings premium that is between the OLS and FE results.

Our within-twin pair analyses of the non-monetary returns suggest that entrepreneurs work longer hours and have more responsibilities at work. On the positive side, they face more often non-monotonous work and can control their working methods.

When we combine these two sets of findings, our results suggest that the returns-to-entrepreneurship puzzle might be two-dimensional, consisting of negative selection into entrepreneurship and of different monetary and nonmonetary returns to entrepreneurship. That is, there appears to be negative selection into entrepreneurship and after entry to entrepreneurship there is a negative monetary impact and evidence of a higher work load, but also of more non-monotonous work and more influence on pace and methods of work. The higher earnings of employees can therefore be interpreted as a compensating differential for monotonous work and the lack of influence.

[4] A number of cross-sectional studies (e.g. Blanchflower and Oswald, 1998; Blanchflower, 2000; Hundley, 2001) and also panel studies (e.g. Benz and Frey, 2004; Taylor, 2004; Kawaguchi, 2008; Andersson, 2008) argue that it may be more satisfying to be self-employed than to work as an employee for an organization. Consistent with this, the desire for independence appears to predict entry into self-employment (e.g. Taylor, 1996, 2004).

[5] The idea that panel data estimators that rely on within-individual variation may fail to capture the longer-term earnings effects has previously been considered in a number of different contexts. For example, Korenman and Neumark (1991) and Isacsson (2007) consider the possibility that they underestimate the long-term income effect of being married.

[6] We use a Finnish twin data set. Because of the high response rate, the data set is as large as the twin data sets in other Nordic countries (Denmark, Norway and Sweden).

[7] The genetic basis of entrepreneurship is not negligible in our data, as it appears that an important share of the variation in the entrepreneurial status is related to genetic factors; see Section 2.2 for further information.

[8] We acknowledge at the outset that twin differencing is not a panacea. For example, twin differencing may not completely remove within-twin variation in unobserved characteristics that correlate with returns to entrepreneurship, leading to an endogeneity problem.

The rest of the paper is organized as follows: In the next section, we study the monetary returns to entrepreneurship. After discussing the identification of entrepreneurial returns with different estimators we describe the data; compare the estimation samples of the traditional panel data estimators and the within-twin estimator; look at the average (log) earnings of those observations used for identification of returns to entrepreneurship with different estimators; and present the estimation results. In the third section we analyze empirically the non-monetary returns to entrepreneurship. Section four concludes.

2. Monetary returns to entrepreneurship

2.1. Econometric framework

We build our empirical analysis of monetary returns to entrepreneurship on the following Mincer-type earnings equation:

$$y_{ijt} = \alpha + \gamma ENT_{ijt} + X'_{ijt}\lambda + \varepsilon_{ijt} \tag{1}$$

where y_{ijt} refers to the (natural logarithm of) earnings of individual i ($i = 1, 2$) from pair (family) j ($j = 1, 2, \ldots, N$) at time t ($t = 1, 2, \ldots, T_{ij}$), ENT_{ijt} is a dummy variable indicating whether a person is an entrepreneur (self-employed) at time t, X_{ijt} is a vector of control variables, and ε_{ijt} is an error term.

The parameter of interest is γ, which measures in percentage terms how much more (or less) entrepreneurs earn, on average, when compared with employees.The error term is assumed to be

$$\varepsilon_{ijt} = \beta A_{ij} + \nu_{ijt} \tag{2}$$

where A_{ij} reflects (time-invariant) unobserved heterogeneity and where ν_{ijt} is an *i.i.d.* random component. In this formulation, A_{ij} is typically interpreted to reflect the unobserved ability of individual i of pair j ("innate ability"), but it could also reflect, e.g., risk aversion, personality traits, and preferences.[9]

Estimating (1)–(2) by standard cross-sectional methods, such as OLS, rests on the assumption that either ability does not affect earnings (i.e., $\beta = 0$) or that conditional on X_{ijt}, ability is uncorrelated with the decision to become an entrepreneur. The standard panel data estimators, such as FD or FE, may also be unreliable, because they tend to pick up a peculiar entry and/or exit phase of the dynamics of the earnings of entrepreneurship.

To see how, consider first the FD-estimator. The variation that this estimator utilizes in identifying γ becomes salient if we first-difference (1) to obtain

$$y_{ijt} - y_{ijt-1} = \gamma\left(ENT_{ijt} - ENT_{ijt-1}\right) + \left(X_{ijt} - X_{ijt-1}\right)'\lambda + \left(\upsilon_{ijt} - \upsilon_{ijt-1}\right) \tag{3}$$

Eq. (3) shows that while FD can remove the ability bias, it uses only certain pairs of observations to identify the parameter of interest (γ): The *last* year of employment prior to entrepreneurship and the *first* year of entrepreneurship on the one hand and the *last* year of entrepreneurship and the *first* year of employment post entrepreneurship on the other hand. This is so because the difference $ENT_{ijt} - ENT_{ijt-1}$ is zero in all the other cases. The direction of the bias depends on how the last and first years of entrepreneurship and employment differ from the average years of entrepreneurship and employment. If there is for example a dip in earnings before a switch to entrepreneurship (similar to Ashenfelter's (1978) dip in the program evaluation literature) and/or a drop in employment income after an exit from entrepreneurship, the FD-estimator would be biased upwards.

The direction of bias is less clear in the case of a FE estimator, as it makes use of variation in the demeaned data:

$$y_{ijt} - \bar{y}_{ij} = \gamma\left(ENT_{ijt} - \overline{ENT}_{ij}\right) + \left(X_{ijt} - \overline{X}_{ij}\right)'\lambda + \left(\upsilon_{ijt} - \overline{\upsilon}_{ij}\right). \tag{4}$$

The benefit of the FE estimator compared with FD is that it uses data on all observations, not just those adjacent to changes in occupation. However, it also relies on the potentially problematic within-individual variation of the data, as $ENT_{ijt} - \overline{ENT}_{ij}$ varies only for those persons who switch their occupation during the sample period. The consistency of the FE-estimator requires that the expected value of the random component of (2) is zero, conditional on ability and the leads and lags of the explanatory variables. This strict exogeneity condition is violated if individuals make career choices based on their expectations of future earnings shocks or if current earnings shocks determine future career choices. Thus, if there is for example a drop in earnings in the last year of entrepreneurship prior to employment and if this drop in part determines the subsequent career choice, the condition is violated.

Twin-differencing can control for unobserved heterogeneity without relying on the dynamic variation in the data (see, e.g., Ashenfelter and Krueger, 1994; Ashenfelter and Rouse, 1998; Bonjour et al., 2003; Isacsson, 2007; Gurrin et al., 2006; Carlin et al., 2005): It removes the ability bias and does not use variation around the entries and exits. The ability bias is removed for identical MZ twins who are "identical at conception".[10] Assuming that $A_{1j} = A_{2j}$ holds for MZ twins, twin differencing (1) and (2) gives

$$y_{1jt} - y_{2jt} = \gamma\left(ENT_{1jt} - ENT_{2jt}\right) + \left(X_{1jt} - X_{2jt}\right)'\lambda + \left(\upsilon_{1jt} - \upsilon_{2jt}\right). \tag{5}$$

This shows that the parameter of interest can be identified from a single cross section through variation in the within pair differences of the variables, calculated for each individual within a twin pair. Standard OLS estimation of (5) is a way to implement the WT estimator.

Three points about the WT-estimator are worth noting: First, the WT-estimator assumes that *conditional on within-twin differences in the observables*, differences in twins' career choices (within pair differences in ENT_{ijt}) represent random variation and thus that the conditional expectation of the differences of the individual earnings shocks is zero within each pair. We acknowledge that this assumption can be problematic (see, e.g., Bound and Solon, 1999; Neumark, 1999), but would like point out already here that, relative to the prior economics literature using twin data, we have at our disposal a rather rich set of control variables that can be used to control for within MZ-pair differences. Moreover, our MZ estimates turn out not to be sensitive to the precise specification of the included control variables. If one accepts the view that unobserved heterogeneity is unlikely to be completely orthogonal to our long vector of control variables, it seems that unobserved heterogeneity may be of lesser concern to us than in some prior studies.

The second point to note about the WT-estimator is that selection is an issue in our twin differenced data to the extent that there remains unobserved heterogeneity that is correlated both with shocks to entrepreneurial earnings and occupational status even after twin differencing and conditioning on our observables. A source of such heterogeneity is the possibility that the twins have been reared apart and, thus, that they do not have similar family and

[9] One can think of it as a sum of the family (pair) effect and the deviation of individual i from the family average.

[10] Identification based on twin differencing is less convincing for DZ twins who share 50% of their genes (assuming random mating). However, using both MZ and DZ twins one gets more observations. For example Black et al. (2007) use twin differencing on a sample of both MZ and DZ twins. In our data, using the combined MZ and DZ sample does not change the point estimates but increases the efficiency of estimation; see Section 2.4 for further discussion.

neighborhood experiences. This, however, is not a great source of concern in the twin data that we use.[11]

The final point about the WT-estimator is that it uses different individuals for identification than the traditional panel data estimators. As noted above, the FD and FE estimators do not use the individuals who do not switch occupation, whereas the WT estimator uses them, as long as the individuals within the twin pair are in different occupations. On the other hand, the WT estimator does not use some of the individuals included in FE or FD, i.e., those twin pairs where the twins switch occupation at the same time but there is no within-twin pair difference in occupation.

2.2. Data and definition of variables

Our data are based on the older Finnish Twin Cohort study. A sample of twins born before 1958 was identified for the large scale survey, which was carried out at three points in time: 1975, 1981, and 1990. Each time, respondents were asked a battery of questions related to their between-twin differences, their medical history, their self-reported experiences at work, their living habits, and their occupation.[12] The surveys cover almost all same sex twin pairs alive in Finland at the time of the survey and born before 1958.[13] The original data contain 11 927 twin pairs and thus 23 854 individuals.

The twin surveys have for this study been matched with the Finnish Longitudinal Employer–Employee Data (FLEED) data base of Statistics Finland. Unlike the twin surveys, FLEED is a register-based annual panel that is a combination of various administrative registers on individuals and firms. The information on individuals in FLEED is based on the Employment Statistics (ES) data base, which includes information on the labor market status of individuals and their background characteristics from different administrative registers. It covers effectively the whole working-age population of Finland, so we can track over time the labor market behavior of the twins who are in the original twin surveys. The FLEED sample available to us covers years from 1988 to 2004.

The main data that we use to estimate the earnings equations consist of those twin pairs for which both individuals can be observed to be alive throughout the FLEED sample period and for which there are no missing data in the key dependent or explanatory variables. To be in line with the previous research on entrepreneurship, we drop farmers (i.e., twin pairs of which at least one is a farmer). We concentrate on male twins to avoid the problem of simultaneously estimating the entrepreneur–employee pay gap, the gender pay gap, and the effects of gender differences in labor supply. In the estimation sample we have 7187 observations of male MZ twins. In robustness analyses we also use 13 724 observations of non-identical (dizygotic, DZ) male twins.

In the FLEED data the definition of an entrepreneur is based on whether the person belongs to the pension system of entrepreneurs. This is a compulsory system for those who satisfy certain requirements (earnings over a minimum limit, entrepreneurship has lasted at least for four months, the person owns alone at least 30% of the firm or at least 50% together with his/her family). The definition therefore refers both to sole trading self-employed individuals as well as to the owners of firms who employ other workers. Based on the register information we define the dummy *Entrepreneurship* (corresponding to ENT_{ijt} in Eqs. (1)–(5)), which equals one if person *i* from pair *j* is an entrepreneur at *t* time and is zero otherwise. We can check the quality of this measure by comparing it to the twin surveys, which have a question on the occupational status.[14] The self-reported entrepreneurial status is highly correlated with the definition in the register data: In 1990, the only year for which this information is available from both data sets, the correlation of the two definitions for MZ males is 0.79 and highly significant ($p<0.0001$).[15]

Measuring entrepreneurs' income is notoriously difficult (e.g. Hamilton, 2000). Therefore, and following Berglann et al. (2011), we employ as comprehensive a measure as possible, which in our FLEED data means that we use the annual sum of wages and salaries, entrepreneurial income, and capital income.[16] These data come from tax registers. We include capital income because entrepreneurs can partially choose whether they take their earnings in the form of salary or capital income, such as dividends. The earnings variable can be computed for 1993–2004 and is deflated with the consumer price index. The dependent variable that we use in the regression analysis is the natural logarithm of earnings (*Ln_earnings*).

The control variables in the earnings equation (X_{ijt}) are drawn from the FLEED data and from the survey. They can be divided into two subsets. The first subset of control variables come from the 1990 survey and include *Age* (in years; we employ a fourth order polynomial), *Height* (in centimeters), *Body mass index*, whether the person had in 1990 experienced *Unemployment less than five years ago*, whether the person had in 1990 experienced *Unemployment more than five years ago*, whether the person was *Lighter than co-twin* (as opposed to being heavier or having same weight) at age 10, and whether the person was a *Current daily smoker* in 1990.[17] Except for *Age*, all of these variables are time invariant over the sample period. Though these variables (except age) drop out in FD and FE estimations, in WT estimations they are included (but age drops out). They allow us to narrow down, relative to many other studies using twin data, the possibility that within-twin differences (not driven by genetics but correlated potentially with the career choices) cause a bias in the results.

In the second subset of controls all variables vary (at least for some individuals) over time. They come from the FLEED data and include *Education years* (based on information on achieved degrees and standard degree times), *Marital status* (=1 if married, =0 otherwise), *Owner of a house or a dwelling* (=1 if owns one, =0 otherwise), *Number of children under age of 7* in the family, *Number of children between 7 and 18 years of age* in the family, *Ownership of taxable wealth* (=1 if

[11] There is a (retrospective) question in the twin surveys that allows us to explore this: It turns out that in our sample, only 3% of the twins have been less than 15 years together and 13% have been together less than 17 years. This suggests that most of the twins in our data have been reared together.

[12] The response rate in the 1975 survey was 89%, whereas in 1981 and 1990, it was 84% and 77%, respectively. For further information and for a detailed description of these surveys, see Kaprio et al. (1979), Kaprio and Koskenvuo (2002), Silventoinen et al. (2004) and www.twinstudy.helsinki.fi.

[13] The 1990 survey was sent only to those twins who were born after 1930. Because we use variables from this survey, it means that the twins in our sample were all less than 60 years old in 1990.

[14] The survey question inquired the respondent's labor market status at the time of the interview. One of the six categories that the respondent was allowed to choose was non-agricultural entrepreneurship.

[15] We can examine the genetic basis of *Entrepreneurship* using the DeFries and Fulker (1985) regression approach, which is a standard behavioral genetics design that measures the importance of genetic heritability and shared environmental factors in generating individual variation in outcomes. The simplest way to proceed is to pool the person-year observations over time and apply the variance decomposition to *Entrepreneurship* in the pooled data. We find that depending on the specification, 25–30% of the variation in *Entrepreneurship* is related to genetic factors, but the role of the common environment is low. To examine this issue in more detail (and, in particular, to take into account the fact that *Entrepreneurship* is binary), we run an alternative set of DeFries and Fulker-regressions using the share of years a person is an entrepreneur in the 1990–2004 sample window as the dependent variable. The estimations show that, depending on the specification, the fraction of the variance explained by genetic heritability varies from 22% to 44%. These exploratory analyses of the genetic basis of entrepreneurship show that the genetic component in entrepreneurship is significant, as found earlier by Nicolaou et al. (2008a).

[16] Wages and salaries include income paid to persons during the year in pay — either in money or benefits in kind. Wages and salaries also include overtime compensation, income received from secondary jobs, and (realized) incentive stock options. Entrepreneurial income includes, for example, income from business activity and business group, and copyright fees. See http://www.stat.fi/meta/kas/index_en.html (accessed 30/1/2008). Capital income includes e.g. dividends, capital gains and interest income.

[17] Differences in the birth weight may have long-term labor market consequences (see, e.g., Black et al., 2007). While imperfect, *Lighter than co-twin* is our proxy for such heterogeneity.

the amount of taxable wealth >0, =0 otherwise), and *Amount of taxable wealth* (=natural logarithm of (1+taxable wealth)).[18] In robustness analysis we use an alternative measure of education, which is a dummy for *High education* (=1 if at least a bachelor's degree, =0 otherwise).

We display the descriptive statistics of these variables in Table A1 in the Appendix, separately for entrepreneurs and employees. Regarding the descriptive statistics on entrepreneurship it should be noted that the probability of entrepreneurship for twins (0.123 in our data, see Panel A of Table A1) is very similar to that in the general population: Using a one-third random sample from the FLEED (which is not restricted to twins), we find that the proportion of entrepreneurs was on average 0.134 in 1993–2004. Comparison of entrepreneurs and employees (Panels B and C of the table) shows that entrepreneurs have on average lower earnings, but they have more wealth. An average entrepreneur has half a year less education and is a year older than an average employee. Entrepreneurs have more often experienced unemployment in the more distant past, but less in the last five years. This may be an indication of past unemployment having driven some of the individuals to entrepreneurship.

2.3. Comparison of estimation samples

Different estimators rely on different parts of the estimation sample for identification. The first difference comes from potentially different samples: The OLS and panel estimators can utilize data on non-twin individuals, yielding much larger samples. The second, often overlooked, difference comes from those twin-observations the different estimators utilize. One way to look at the differences is to compare the frequency and overlap of those MZ twin observations that yield identification in the sample. This comparison shows that 7.9% of the observations are from individuals who switch occupation at least once during our observation period. It is this subsample whose observations are used for identification with the traditional panel data estimators. The FD estimator only uses a (particular) subset of these observations, the FE estimator all these observations. In comparison, 14.1% of our observations are such that the twins are in different occupations in a given year; it is these observations that the within-twin estimator uses for identification. The correlation between the dummies indicating inclusion in these samples is only 0.2 and the samples share only 569 observations, suggesting that very different observations and (observations on different) individuals deliver identification for the different estimators.

Turning then to the observations delivering identification for the traditional panel data estimators, we calculated the means of *Ln_earnings* for 1) the observations that are a year before individual *i* from pair *j* changes from being an employee to becoming an entrepreneur; 2) the first year of entrepreneurship; 3) the last year of entrepreneurship; 4) the first year after entrepreneurship; 5) and averages for the periods prior to entrepreneurship, of entrepreneurship, and after entrepreneurship, always excluding the years of occupational change, and the years just before and just after.

The FD estimator uses just the year of occupational change and the year before. As row 1 of Table 1 shows, the mean *Ln_earnings* in the year before an individual became an entrepreneur is 10.13, while the mean for the first year of entrepreneurship is 10.37, suggesting a return of 10.37 − 10.13 = 0.24. Using the last year of entrepreneurship (average 10.11) and first year after entrepreneurship (9.97) suggest a return of 0.14.

We then compare the last year before entrepreneurship and the first year after entrepreneurship to the average *Ln_earnings* of the same individuals in other years of paid employment. The difference between *Ln_earnings* in the last year before entrepreneurship and the average *Ln_earnings* prior to entrepreneurship (excluding the last year) is −0.10, i.e., these individuals earn less in the last year before becoming entrepreneurs than they have earned in prior years. The difference is not statistically significant, but that is probably mainly due to the small number of observations we have for the earnings in the last year before entrepreneurship. The first year of paid employment after entrepreneurship yields earnings that are lower than the average earnings in paid employment both before (−0.26) and after (−0.41) entrepreneurship. These differences are large, too. Even if there were no systematic differences in the earnings of the entrepreneurs in their first, last, and other years, these differences would suggest that the FD estimates of returns to entrepreneurship should be upwards biased.

A comparison of the income of entrepreneurs in different years reveals some systematic differences, however. The average *Ln_earnings* in the first year of entrepreneurship is 10.37, compared with average of 10.07 in all other years of entrepreneurship bar the first and the last.[19] Again, it seems that the reason for the statistical insignificance is the small number of observations. In contrast, the earnings in the last year of entrepreneurship, at 10.11, are close to the average of all years in entrepreneurship bar the first and the last.[20]

Taken together, these differences suggest that the FD estimator would yield an upward biased estimate of returns to entrepreneurship because of three reasons: 1) the earnings in the last year of paid employment prior to becoming an entrepreneur are lower than in other years of paid employment (like in Ashenfelter's dip); 2) the earnings in the first year of paid employment after entrepreneurship are lower than in other years of paid employment; and 3) the earnings in the first year of entrepreneurship are higher than in other years of entrepreneurship.[21]

The FE estimator uses all the observations of the individuals that change occupations. However, the strict exogeneity assumption on which the FE estimator's consistency rests requires in our context that individuals change occupation neither because of this year's nor because of future years' earnings shocks. The fact that the earnings of those individuals who change occupations are lower in the last year before entrepreneurship than in the other years of paid employment prior to entrepreneurship suggests that the strict exogeneity assumption may be violated. Similarly, the fact that the earnings in the first year of entrepreneurship are higher than in the other years of entrepreneurship suggests that knowledge of this earnings shock may have contributed to the fact that these individuals changed jobs. This pattern would be another violation of strict exogeneity.

2.4. Empirical analysis

We present our results for the earnings regressions in Table 2. Unless otherwise indicated, these results are based on estimations that include year dummies (OLS, FD, FE) and both the register and survey based subsets of control variables. Standard errors that are robust to heteroscedasticity are displayed in square brackets. More conservative

[18] Our results are robust to not using wealth as a control variable. We include wealth to control for the possibility that a substantial amount of some entrepreneurs' wealth may be invested in their business. Such individuals no longer receive capital income that this wealth would generate, had it been invested in other assets.

[19] In an interesting paper, Tergiman (2011) reports, using the 2001 Survey of Income and Program Participation data, that entrepreneurs earn more (in hourly wages) relative to employees both early and late in during their entrepreneurial spell. The figures in Table 1 are not directly comparable, but do suggest that the earnings profile for entrepreneurs has a U-shape.

[20] It also is of interest to note that the descriptive statistics of Table 1 suggest that those individuals leaving entrepreneurship earn substantially more after than during entrepreneurship, bar the first year after entrepreneurship.

[21] The data also confirm the previous findings that the variance of earnings is higher for entrepreneurs than for employees. The standard deviation of *Ln_earnings* during entrepreneurship (excluding the first and last year) is 1.02, that of employment income prior to entrepreneurship 0.72, and that of employment income after entrepreneurship 0.54.

Table 1
Comparison of wages during different occupational periods.

			Last year before entr.	First year as entr.	Last year as entr.	First year after entr.
		Ln_earnings	10.13	10.37	10.11	9.97
	Ln_earnings	Obs.	23	24	25	30
Before entr.	10.23	99	-0.10	0.14	-0.12	-0.26
During entr.	10.07	192	0.07	0.31	0.05	-0.09
After entr.	10.38	212	-0.25*	-0.01	-0.27*	-0.41***

Notes: The reported numbers are i) the natural logarithm of earnings (*Ln_earnings*), ii) the number of observations (Obs.) and iii) in the area bordered from above and left by the dashed lines, the difference between the column variable and the row variable. "Before entr." refers to the average *Ln_earnings* of all observations prior to an entrepreneurial spell, bar the last period before the spell; "During entr." refers to the average of *Ln_earnings* of all entrepreneurship observations, bar the first and last periods before and after an entrepreneurial spell; "After entr." refers to the average of *Ln_earnings* of all observations after an entrepreneurial spell, bar the first period after the spell.
***1% significance level.
*10% significance level.

standard errors that allow for clustering are in parentheses. They are clustered by twin pair in OLS and WT, allowing for within-individual and within-twin correlation. In FD and FE the standard errors are clustered by individual.

The table consists of several panels: In Panel A we present our main results, while the other panels examine the internal validity of our findings, i.e., the robustness of the results to dropping the controls, using various subsamples, and dropping outliers. In all panels, column (1) contains the OLS estimates produced by pooling all the person-year observations over the whole FLEED sample period 1993–2004, whereas columns (2)–(4) display the FD, FE and the WT estimates, respectively.

The first thing to note about the estimation results is the ranking of the estimates that they imply: $\hat{\gamma}_{OLS}<\hat{\gamma}_{WT}<\hat{\gamma}_{FE}<\hat{\gamma}_{FD}$. The OLS estimate of *Entrepreneurship*, −0.44, is the smallest and also highly significant statistically. It is followed by the WT estimate, −0.23, which is significant at the 12% level if the clustered standard errors are used (and at the 1% level if the heteroscedastic-robust standard errors are used). The FE and FD estimates of Panel A are larger, but insignificant.[22]

Three aspects of these results are worth being given emphasis: First, the ranking of the estimates is consistent with the interpretation that there is unobserved heterogeneity, which pushes the OLS estimates downwards, and that using within-individual variation to control for it (i.e., using the FD and FE estimators) is hampered by the dynamics of entrepreneurial entry and exit. The second aspect worth emphasizing is that the ranking of the estimates is robust to various modeling choices. Third, as we demonstrate below, there is a reason to think that our results apply to the general Finnish working population.

2.5. Interpretation of the estimate ranking

To start with, the apparent downward bias in the OLS estimate is consistent with unobserved heterogeneity that, conditional on the observable characteristics (e.g., education), leads to negative selection into entrepreneurship. In support of this, Hyytinen and Rouvinen (2008) report, based on an analysis of the European Community Household Panel, that there appears to be negative selection into entrepreneurship in Europe. Uusitalo's (2001) results shed some light on how this selection may come about: He reports (using Finnish data on males) that entrepreneurs obtained lower verbal but higher math scores in the standardized tests administered to them at the beginning of their military service. Entrepreneurs scored higher in "leadership" and "dynamism" but lower in "cautiousness".[23] On the other hand, an obvious explanation for the clearly less negative FE and positive FD estimates is that they use for identification either unusually good years in terms of earnings at the beginning and end of the entrepreneurial spells or, alternatively, unusually bad years just prior to or after the spells. That is, the FD and FE estimators may be biased upwards, because they do not adjust for the possibility that there is a dip in the employment income before a switch to entrepreneurship (like in Ashenfelter, 1978) and, again, after the entrepreneurial spell. Such a temporal pattern could for example arise if individuals prepare for entrepreneurial entry (exit) by devoting their time and effort to it while still employed (self-employed). To the extent that mechanisms of this type are more intensive close to the eventual occupational transition, this also explains why the FD estimate is larger than the FE estimate.

Although our interest is on the returns to entrepreneurship, inclusion of the *Education years* to the models of Table 2 allows us to assess the returns to education and, in particular, to compare how the estimated returns vary between the various estimators. For brevity, we discuss the results without presenting them in the table. Because the years of education of a person can typically only increase (i.e., there is no potentially complicated dynamics underlying this explanatory variable) and because selection to education may well be positive, we have no reason to expect that the coefficients of the various estimators would be ranked similarly to those of *Entrepreneurship*-variable. This is indeed what we found. The returns to one additional year of education were from 0.06 to 0.07 in OLS estimation. The FD results were of the same order of magnitude, but varied more from one specification to another. The FE results were in most cases slightly lower and sometimes insignificant.[24] The WT estimates were the lowest, from 0.02 to 0.025, and mostly statistically insignificant. These numbers and e.g. the differences between the OLS and WT estimates are broadly in line with the earlier (see, e.g., Taubman, 1976; Bonjour et al., 2003), and more recent literature (see, e.g., Zhang et al., 2007; Li et al., 2012, and the references therein) that uses twin data to estimate the returns to education. These results suggest that the ranking of the estimates for *Entrepreneurship*-variable is driven by the peculiar dynamics of events that surround entrepreneurial entry and exit.

2.6. Internal validity and robustness

We explore the internal validity and robustness of our results in Panels B to H of Table 2. Using both MZ and DZ male twins in Panel B almost triples the sample size. For the WT estimator, there is a trade-off between the MZ and combined MZ and DZ samples between how well twin-differencing purges the ability bias on the one hand and the potentially small size of the (sub)sample from which the key parameters of interest are identified on the other hand. The results show that the WT estimate stays close to what it was in the MZ sample, but it is now clearly significant. It seems that the benefit of the larger sample size outweighs the cost of less effective purging of the unobservables. There are more changes in the other estimates (e.g., the OLS estimate becomes less negative, whereas the FD and FE estimates decrease), but the ranking of the estimators does not change.

[22] The set of controls is clearly significant in OLS, FD, and FE estimations, but not in WT estimation. This indicates that the observable characteristics of the twins are relatively similar.

[23] These differences between individuals selecting into (or out of) entrepreneurship may be behind the observed downward bias in the OLS estimates. Uusitalo also finds that entrepreneurship "runs in the family", i.e., both parents' self-employment increases the likelihood of an individual becoming an entrepreneur. Our within-twin results are robust to such selection (as are FD and FE estimates).

[24] Since even the youngest individuals are in their mid-30s in the beginning of the panel data, there are relatively few changes in their education and therefore the panel methods identify the returns to education from a small number of observations.

Table 2
Estimation of monetary returns to entrepreneurship.

		OLS	FD	FE	WT
Panel A: baseline model		−0.444	0.133	−0.016	−0.230
		(0.115)***	(0.170)	(0.150)	(0.146)+
		[0.049]***	[0.175]	[0.150]	[0.072]***
	Obs.	6546	5614	6546	3292
Panel B: MZ and DZ twins		−0.308	−0.076	−0.172	−0.247
		(0.060)***	(0.088)	(0.087)**	(0.077)***
		[0.027]***	[0.086]	[0.087]**	[0.037]***
	Obs.	18867	16109	18867	9497
Panel C: no controls		−0.333	0.121	−0.016	−0.208
		(0.114)***	(0.166)	(0.148)	(0.152)
		[0.048]***	[0.171]	[0.148]	[0.072]***
	Obs.	6546	5620	6546	3292
Panel D: dummy for *High education*		−0.470	0.133	−0.013	−0.243
		(0.114)***	(0.170)	(0.150)	(0.145)*
		[0.048]***	[0.175]	[0.150]	[0.072]***
	Obs.	6546	5614	6546	3292
Panel E: 6 or more working months		−0.466	0.163	−0.054	−0.239
		(0.113)***	(0.189)	(0.163)	(0.147)+
		[0.048]***	[0.194]	[0.163]	[0.073]***
	Obs.	6377	5512	6377	3212
Panel F: birth year after 1947		−0.490	0.093	−0.089	−0.311
		(0.167)***	(0.222)	(0.180)	(0.189)+
		[0.070]***	[0.229]	[0.180]	[0.092]***
	Obs.	4142	3642	4142	2084
Panel G: observations with high Cook's distance dropped		−0.181	0.104	0.105	−0.074
		(0.062)***	(0.042)**	(0.061)*	(0.076)
		[0.024]***	[0.042]**	[0.061]*	[0.035]**
	Obs.	6240	5416	6252	3120
Panel H: observations with high DFBETA dropped		−0.320	0.136	−0.011	−0.104
		(0.035)***	(0.019)***	(0.031)	(0.039)***
		[0.018]***	[0.020]***	[0.031]	[0.024]***
	Obs.	6055	5577	6303	3078

Notes: The dependent variable is *Ln_earnings* and the coefficient and its std. errors are for *Entrepreneur*-dummy. Standard errors in parentheses are clustered by twin pairs in OLS and WT, and by individual in FD and FE. Robust standard errors (without clustering) in brackets. Significance level: ***1%, **5%, *10%, +15%. The controls are a fourth order polynomial in *Age*, *Height*, *Body mass index*, *Unemployment less than five years ago*, *Unemployment more than five years ago*, *Lighter than co-twin*, *Current daily smoker*, *Marital status*, *Owner of a house or a dwelling*, *Number of children under age of 7*, *Number of children between 7 and 18 years of age*, *Ownership of taxable wealth*, and *Amount of taxable wealth*. In WT estimations, the polynomial of *Age* is not included. In FD and FE time-invariant variables measured in 1990 are not included. Time dummies are included except in WT. *Education years* is included except in Panel D.

In the other robustness checks we use the MZ sample. In Panel C of Table 2 we have dropped the controls (except for the years of education and year dummies). This has relatively little impact on the results, with the exception of OLS, where the estimate increases. In Panel D we replace *Education years* with a *High education*-dummy. This does not change the conclusions on the returns to entrepreneurship. In Panel E we restrict the sample to those who have at least six working months during the year. This drops individuals with less attachment to the labor market, e.g. those with long unemployment spells or non-employment between job switches. This does not drop the sample size much and the results remain intact.[25] In Panel F we change the sample to include only those individuals who were in their prime working age during our observation period. These we take to be individuals born between 1948 and 1957. The motivation is that those born earlier were approaching or reached the retirement age towards the end of our sample period. All of the estimates drop relative to baseline results in Panel A, with the WT estimate experiencing the biggest drop to −0.31 (significant at 10.2% level when the clustered standard errors are used).

For the two final robustness checks we drop observations which can be identified as potential outliers. In Panel G we use Cook's distance to identify the outliers separately in each of the estimations.[26] There is a modest drop in the sample size, but this has a surprisingly large effect on the estimates. Especially the OLS and WT estimates increase: The latter is still negative, but becomes insignificant when the clustered standard errors are used (but is still significant when heteroscedastic-robust standard errors are used). Also the FE and FD estimates change, but importantly, the ranking of the WT estimator relative to the panel data and OLS estimators is preserved also in this case. Finally, in Panel H we examine the effect of dropping observations that have the biggest impact on the coefficient of *Entrepreneurship*. These observations are identified in each of the estimations with the help of the so-called DFBETA measure.[27] Dropping observations with high DFBETA decreases the number of observations a bit more than the use of Cook's distance in Panel G. There is a clear increase (relative to the baseline results) in the OLS and WT estimates, but the ranking of the estimates stays the same. Moreover, the WT estimate of the returns to entrepreneurship, −0.10, is now again clearly significant.[28]

In sum, the ranking of the estimates, $\hat{\gamma}_{OLS}<\hat{\gamma}_{WT}<\hat{\gamma}_{FE}<\hat{\gamma}_{FD}$, appears to be quite robust, as is the finding that the monetary returns to entrepreneurship are negative when one controls for unobserved heterogeneity using within-twin variation instead of temporal within-individual variation.

2.7. External validity

Our analysis has so far used data on a sample of twins. This naturally raises the question of external validity of our findings, i.e., the extent to which the twin results can be generalized to the wider population. To address this question, we compare the twin results to a set of corresponding regressions that make use of a large (one-third), representative random sample of Finns (from the FLEED data of Statistics Finland). This non-twin sample allows us to re-estimate some of the OLS, FE and FD models that we report in Table 2. We use for these comparisons the same (age) cohorts and data period from the non-twin data as we have in the twin data. It turns out, for example, that the OLS, FD and FE coefficients of *Entrepreneurship*-dummy and *Education years*, which are the two main variables of interest to us, are very similar to those we obtain using the twin sample: In the representative non-twin sample, the coefficients are −0.331 (OLS), −0.106 (FD) and −0.200 (FE) for *Entrepreneurship* and 0.066 (OLS), 0.016 (FD) and 0.023 (FE) for *Education years*.[29] While the match is not perfect, these numbers suggest that our twin findings can be generalized to the wider population.

So far we have focused on monetary returns to entrepreneurship. However, because the prior literature strongly suggests that non-monetary returns also matter, we complement our main analysis by studying briefly the non-earnings aspects of the returns to entrepreneurship.

[25] A remaining concern that we cannot fully address here is that entrepreneurs may work longer hours than employees and that this introduces a bias in the estimates of the monetary returns. We cannot address this in detail, because the register-based earnings data in the FLEED are annual and because, unfortunately, there is no data on the annual total hours worked. See the conclusions for further discussion.

[26] Cook's distance measure is calculated for each observation using the estimates in Panel A and those observations for which the measure is greater than 4/*N* (where *N* is the number of observations) are dropped.

[27] Also this measure is calculated for each observation using the estimates in Panel A. Observations for which the absolute value of DFBETA exceeds 2 divided by the square root of *N* are dropped.

[28] We estimated the WT models in panels C to H of Table 2 also using the combined MZ and DZ twin sample. The results were close to those obtained with the MZ data and clearly significant even when clustered standard errors are used. This supports the conclusion from the comparison of Panels A and B that the low significance of some of the WT results in Table 2 is driven by the small sample size.

[29] All these coefficients are highly significant statistically.

3. Non-monetary returns to entrepreneurship

3.1. Econometric framework

To provide a framework for the analysis of non-monetary returns, let w_{ijt} be 1 if there is a positive non-earnings aspect of work for an individual i from pair j at time t, and 0 otherwise. We assume that there is a latent variable, w^*_{ijt}, such that $w_{ijt}=1$ if $w^*_{ijt}>0$ and $w_{ijt}=0$ if $w^*_{ijt}\leq 0$. The latent variable follows

$$w^*_{ijt} = \alpha + \gamma ENT_{ijt} + X'_{ijt}\lambda + \varepsilon_{ijt}. \tag{6}$$

where $\varepsilon_{ijt}=\beta A_{ij}+\nu_{ijt}$, and ν_{ijt} has a standard logistic distribution.

We can now apply the method of twin differencing to this model: Take one cross-section and consider individual ($i=1,2$) and twin pair ($j=1,\ldots,N$) to be the dimensions of the data. Assuming, as before, that $A_{1j}=A_{2j}$ for the MZ twins in (6), it is possible to apply the conditional logit (fixed-effects logit) estimator to control for the unobserved heterogeneity due to A_{ij} (e.g., Chamberlain, 1980; Wooldridge, 2002). However, unlike in the standard panel data setting which uses within-individual variation over time, our estimation uses the twin pairs in which only one of the twins has the particular positive non-pecuniary aspect. The explanatory variables are within-twin pair differences and, as a consequence, the γ parameter is identified through within-twin pair variation in entrepreneurship. We call this "within-twin logit" (WT-Logit) estimator.

3.2. Data and definition of variables

To estimate the non-monetary returns to entrepreneurship, we use only the data from the three surveys of the older Finnish Twin Cohort study carried out in 1975, 1981, and 1990. The surveys provide us with variables that in different ways describe aspects of life which entrepreneurship could potentially influence. Here we concentrate on work-related aspects.[30] The questions are framed in different ways: Some of the answers are binary variables, some counts, some on a Likert scale, and some continuous. To be able to use a common model for all of the survey questions, we have transformed the variables to binary ones, with 1 indicating a positive outcome or aspect. We use the WT-Logit estimator, pooling the survey waves, but even one wave (i.e., a cross-section) is sufficient to produce the WT-Logit estimates.

We report results on the following dependent variables: *Non-monotonous work* (=1 if work is not very monotonous; best two cases in a four-point scale, =0 otherwise), *Working at most 40 hours* (=1 if average weekly working time at most 40 h, =0 otherwise), *Influence on pace of work* (=1 if able to choose work pace; best two cases in a three-point scale, =0 otherwise), *Influence on working methods* (=1 if can influence working methods; best two cases in a three-point scale, =0 otherwise), and *No significant increase in responsibilities at work* (=1 if no increase, =0 otherwise). The number of observations available for each dependent variable varies because of non-response, because the 1990 survey was not sent to those born before 1930, and because most of the questions from which we derive our indicators of non-monetary earnings have been asked in only one of the survey waves.[31]

The information on entrepreneurship is based on the survey respondents' answers to a question about their occupational status. There are 4686 male MZ twin pair observations in the three surveys for which this information is available for both twins. Based on these data, we let the dummy for *Entrepreneurship* be equal to one if person i from pair j is an entrepreneur at time t and is zero otherwise.

The controls used in the analysis of non-monetary earnings are drawn from the surveys. They are *Height* (in the survey year, in centimeters), *Body mass index* (in the survey year), *Education years* (based on standard degree times),[32] and an indicator for *Marital status* (=1 for married or cohabiting during the survey year, =0 otherwise). The descriptive statistics of these variables can be found in Table A2 in the Appendix.[33] Note that we do not have to explicitly control for age, family, cohort or time effects, because they are conditioned out by the twin differencing.

3.3. Empirical analysis

Table 3 presents the results for the WT-Logit estimations (Panel A) and, for comparison, standard pooled Logit results (Panel B).[34] In each case we present only the coefficients of *Entrepreneurship* and *Education years*. Since the dependent variables are coded so that 1 indicates a positive aspect, a positive coefficient is suggestive of non-monetary returns to entrepreneurship.

WT-Logit estimations show that entrepreneurs work more often more than 40 h than employees and that the probability of no significant increase in responsibilities at work is lower for them. At the same time, the work of entrepreneurs is more often non-monotonous and they have more control on the methods and pace of work. [35] The table also shows the relationship of education years to the work aspects. Education years have a clear positive relationship with non-monotonous work. Education is also related to more influence on the pace of work and increases in responsibilities at work, but these relationships are statistically less significant. The pooled Logit results are qualitatively similar but statistically stronger and the parameter estimates are larger in absolute value. We cannot present a systematic comparison between the traditional FE-Logit and WT-Logit estimators because *Non-monotonous work* is the only one of the dependent variables that is available in more than one survey wave. For this variable, the FE-Logit estimates for the coefficients of

[30] In addition to the set of work-related non-monetary aspects, we initially experimented with a wider set of variables. These included various negative shocks in life, like conflicts with spouse or divorce, positive changes in life, smoking, drinking, use of medicines, and various illnesses. Using the WT-Logit estimator that controls for unobserved heterogeneity by twin-differencing we did not find statistically significant differences between entrepreneurs and employees.

[31] The question on monotonous work was in all three surveys, the questions on work pace and increase in responsibilities were in the 1981 survey, and the questions on working time and working methods in the 1990 survey.

[32] Unlike in the earlier analysis on monetary returns that uses the FLEED and refers to 1993–2004, in this section *Education years* also includes studies that have not led to a degree. The reason for this is that here information on education comes from the 1975, 1981 and 1990 twin surveys and not from the register of completed degrees. The respondents could answer for example that they had high school and some years of university. It appears that not all individuals in our data had completed their education at the time of these surveys.

[33] Some of the characteristics of the individuals in the surveys differ from those in the register data. During the survey years the individuals were younger than during the period from which the register data are available and therefore had lower level of education and were less likely to be entrepreneurs. (The age variable is not used in the estimations, but is included in the descriptive statistics to allow comparison of the data sets.) Those who already were entrepreneurs during the surveys were on average older than the employees.

[34] Because of the conditioning in the WT-Logit estimation, only those twin pairs are effectively used in each estimation for whom the dependent variable has variation within a twin pair. If, for example, both twins of a particular pair always have non-monotonous work, the pair is not included in the estimation. As a result, the effective estimation samples reported in the estimation table are (much) smaller than the initial pooled survey sample of male MZ twins.

[35] As a robustness check we estimated fixed effects ordered Logit models for those dependent variables of ours for which the answers to the original survey question were given on an ordered scale. We followed Ferrer-i-Carbonell and Frijters (2004) who suggested that the model can be estimated as a conditional Logit model when the ordered data are collapsed to binary data with unit (in our case twin pair–year) specific thresholds. The recording of observations to "high" and "low" values is based on comparison of individuals to the twin pair's average answers in a given survey wave. The results produced by this alternative method of estimation were similar to those reported in Table 3 and are not reported.

Table 3
Estimation of non-monetary returns to entrepreneurship.

	Non-monotonous work	Working time at most 40 h	Influence on pace of work	Influence on working methods	No significant increase in work responsibilities
Panel A: WT-Logit estimations					
Entrepreneur	0.876	−1.231	0.534	0.884	−0.859
	(0.498)*	(0.471)***	(0.323)*	(0.526)*	(0.408)**
Education years	0.296	−0.019	0.127	0.118	−0.114
	(0.100)***	(0.120)	(0.086)+	(0.119)	(0.074)+
Obs.	866	310	672	304	536
Panel B: Logit estimations					
Entrepreneur	1.161	−0.822	0.666	1.094	−0.439
	(0.231)***	(0.213)***	(0.163)***	(0.191)***	(0.169)***
Education years	0.158	0.087	0.059	0.115	−0.123
	(0.025)***	(0.022)***	(0.018)***	(0.023)***	(0.018)***
Obs.	5511	1234	2089	1235	1995

Notes: The dependent variables are based on survey questions that describe aspects of working life. They are binary variables, coded so that one always indicates a positive outcome or aspect. Standard errors are clustered by twin-pairs. Significance level: ***1%, **5%, *10%, +15%. The controls are *Height*, *Body mass index*, and *Marital status*.

Entrepreneurship and *Education years* were 0.568 (std. error 0.392) and 0.388 (std. error 0.197), respectively.

4. Conclusions

Estimating returns to entrepreneurship is challenging. While it is well-known that controlling reliably for unobserved heterogeneity (e.g. ability, preferences such as risk-aversion, traits, family effects) is very difficult with cross-section data, much less attention has been paid to how entry and exit dynamics affect the standard panel data estimates of the returns to entrepreneurship. Both the FD and FE estimators rely on within-individual variation. The former relies on the potentially peculiar years round occupational shifts, and the latter assumes (in the form of the strict exogeneity assumption) that contemporaneous and future shocks to income do not affect occupational decisions.

We suggest an alternative approach by focusing on twin data. Twin differencing, while not trouble-free, avoids some of the pitfalls buried in conventional estimators used so far. We apply the WT estimator together with the OLS, FD and FE estimators to a large panel data on Finnish male twins and find that the estimates of monetary returns can be ranked in a particular way, i.e. that $\hat{\gamma}_{OLS}<\hat{\gamma}_{WT}<\hat{\gamma}_{FE}<\hat{\gamma}_{FD}$. This ranking is robust to a number of sensitivity checks and consistent with the interpretation that there is unobserved heterogeneity (pushing the OLS estimates downwards) and that using within-individual variation to control for it is hampered by the dynamics of entrepreneurial entry and exit. It thus seems that the FD and FE estimators may be biased upwards, because they do not adjust for the possibility that there is a dip in employment income before a switch to entrepreneurship – as in Ashenfelter's (1978) famous program evaluation example – and, again, after the entrepreneurial spell. The WT estimator, on the other hand, suffers neither from unobserved heterogeneity nor from the dips and suggests a negative earnings premium for entrepreneurs.

When we combine this finding with our within-twin pair analysis of non-monetary returns, our results suggest that the returns-to-entrepreneurship puzzle might be two dimensional, consisting of negative selection into entrepreneurship and of different monetary and nonmonetary returns to entrepreneurship. It seems, in particular, that there is negative selection into entrepreneurship and that after entry to entrepreneurship, there is a negative monetary impact and evidence of a higher work load, but non-monotonous work and more influence on pace and methods of work. The higher earnings of employees can therefore be interpreted as a compensating differential for monotonous work and the lack of influence.

We conclude our paper with three caveats. First, like the FD and FE estimators, the WT estimator relies on differencing (though the dimension is different). The WT estimates may therefore be biased towards zero due to measurement error (Griliches, 1979; Ashenfelter and Krueger, 1994). If they are, the negative monetary returns to entrepreneurship documented in this paper should probably be interpreted as an upper bound estimate. Second, we have estimated the monetary returns to entrepreneurship using earnings measures that are derived from tax registers. A concern that we cannot systematically address here is the extent to which Finnish entrepreneurs are more able than employees to underreport their earnings for tax purposes. If there is a significant difference, it could potentially bias the estimates of monetary returns downwards. While the Finnish tax system and income records are advanced and comprehensive by international standards (and thus probably more reliable than in many other countries), this possibility should not be overlooked, as there is some evidence which suggests that self-employment income is somewhat underreported in Finland (Johansson, 2005). Seen from this perspective, the negative monetary returns to entrepreneurship documented in this paper should be interpreted as a lower bound estimate. Third, our analysis of monetary returns cannot fully address the possibility that entrepreneurs work longer hours than employees. This is not necessarily a major cause of concern, because this difference introduces a positive bias in the estimates of the monetary returns (i.e., it makes all the estimates *less* negative). This means that if anything, our qualitative results on the monetary returns being negative would be strengthened, if we could use, e.g., earnings per hour as the response variable.

A more subtle issue is that the above concerns — i.e., measurement error, differences in the hours worked and variation in the entrepreneurs' and employees' ability to underreport income — bias the OLS, FD, FE and WT estimates differently. It is very difficult to determine conclusively how large such effects might be.[36]

The returns-to-entrepreneurship puzzle is nuanced and multidimensional and, therefore hard to solve. However, our analysis shows that the use of new types of data sets, like twins data, and comparison of different estimators help in clarifying the empirical issues and in understanding the dynamics and economic mechanisms that generate the puzzle.

[36] However, it seems that for e.g. the differences in the hours worked to change the relative ranking of the panel data and WT estimates, what is needed is that the covariance between the hours worked and the entrepreneurial status in the sub-sample of twins in which one of the twins is an employee and the other is an entrepreneur should be much higher than the covariance between (the change in) the hours worked and (the change in) the entrepreneurial status in the first-differenced (FD) or mean-differenced (FE) data. If a twin who becomes an entrepreneur increases his hours by a certain amount and if this amount is approximately the same as the difference in the hours worked between him (when entrepreneur) and his employed co-twin, the difference between the two covariances should be small.

Appendix A. Descriptive statistics

Table A1
Descriptive statistics: Data for the analysis of monetary returns.

Variable	Obs.	Mean	Std. Dev.	Min	Max
Panel A: All persons					
Entrepreneur (dummy)	6546	0.123	0.729	0	1
Panel B: Entrepreneurs					
Ln_earnings	805	9.892	1.343	1.339	12.728
Education years	805	12.035	2.308	9	18
High education-dummy	805	0.349	0.478	0	1
Age	805	49.588	6.067	36	64
Height	805	176.537	6.058	160	193
Body mass index	805	25.233	3.011	18.513	36.228
Unemployment less than five years ago	805	0.006	0.079	0	1
Unemployment more than five years ago	805	0.052	0.223	0	1
Lighter than co-twin	805	0.107	0.309	0	1
Current daily smoker	805	0.581	0.494	0	1
Marital status	805	0.776	0.417	0	1
Owner of a house or a dwelling	805	0.932	0.252	0	1
Number of children under age of 7	805	0.128	0.454	0	4
Number of children between 7 and 18 years of age	805	0.648	1.015	0	7
Ownership of taxable wealth	805	0.839	0.368	0	1
Amount of taxable wealth	805	8.811	4.016	0	13.368
Panel C: Employees					
Ln_earnings	5741	10.265	0.580	5.455	13.976
Education years	5741	12.565	3.049	9	21
High education-dummy	5741	0.360	0.480	0	1
Age	5741	48.692	5.725	36	65
Height	5741	176.458	6.271	160	196
Body mass index	5741	24.63	3.082	16.541	37.551
Unemployment less than five years ago	5741	0.053	0.225	0	1
Unemployment more than five years ago	5741	0.034	0.182	0	1
Lighter than co-twin	5741	0.144	0.351	0	1
Current daily smoker	5741	0.622	0.485	0	1
Marital status	5741	0.821	0.383	0	1
Owner of a house or a dwelling	5741	0.854	0.353	0	1
Number of children under age of 7	5741	0.135	0.442	0	4
Number of children between 7 and 18 years of age	5741	0.616	0.916	0	6
Ownership of taxable wealth	5741	0.654	0.476	0	1
Amount of taxable wealth	5741	6.402	4.758	0	14.900

Notes: This table displays the descriptive statistics for the data used to estimate monetary returns to entrepreneurship. The sources of the data are the 1990 wave of the Finnish Twin Cohort survey and the Finnish Longitudinal Employer–Employee Data of Statistics Finland.

Table A2
Descriptive statistics: Data for the analysis of non-monetary returns.

Variable	Obs	Mean	Std. Dev.	Min	Max
Panel A: All persons					
Entrepreneur (dummy)	4448	0.100	0.300	0	1
Panel B: Entrepreneurs					
Education years	445	8.580	2.358	6	16
Age	445	40.343	9.585	18	63
Height	445	175.665	5.883	154	192
Body mass index	445	25.119	2.872	16.896	36.228
Marital status	445	0.840	0.367	0	1
Non-monotonous work	442	0.943	0.231	0	1
Working at most 40 h	132	0.197	0.399	0	1
Influence on pace of work	163	0.681	0.468	0	1
Influence on working methods	133	0.820	0.386	0	1
No significant increase in responsibilities at work	159	0.579	0.495	0	1
Panel C: Employees					
Education years	4003	9.051	2.799	6	16
Age	4003	35.727	9.882	18	64
Height	4003	175.405	6.242	152	196
Body mass index	4003	24.116	2.877	16.412	38.398
Marital status	4003	0.723	0.448	0	1
Non-monotonous work	3964	0.839	0.368	0	1
Working at most 40 h	726	0.375	0.484	0	1
Influence on pace of work	1562	0.529	0.499	0	1
Influence on working methods	726	0.559	0.497	0	1
No significant increase in responsibilities at work	1495	0.636	0.481	0	1

Notes: This table displays the descriptive statistics for the MZ twin survey data used to estimate non-monetary returns to entrepreneurship. The table covers observations for which all of the control variables and at least one of the dependent variables was available. The sources of the data are the three waves (1975, 1981, 1990) of the Finnish Twin Cohort survey. Due to non-response and changes in the survey questions over the survey waves, the number of observations varies from question to question.

References

Andersson, Pernilla, 2008. Happiness and health: well-being among the self-employed. Journal of Socio-Economics 37 (1), 213–236.

Ashenfelter, O., 1978. Estimating the effect of training programs on earnings. The Review of Economics and Statistics 6 (1), 47–57.

Ashenfelter, Orley, Krueger, Alan, 1994. Estimates of the economic return to schooling from a new sample of twins. American Economic Review 84 (5), 1157–1173.

Ashenfelter, Orley, Rouse, Cecilia, 1998. Income, schooling and ability: evidence from a new sample of identical twins. Quarterly Journal of Economics 113 (1), 253–284.

Åstebro, Thomas, 2011. The returns to entrepreneurship. In: Cummings, Douglas (Ed.), Oxford Handbook of Entrepreneurial Finance. Oxford University Press, pp. 45–108.

Åstebro, Thomas, Chen, Jing, Thompson, Peter, 2011. Stars and misfits: self-employment and labor market frictions. Management Science 57 (11), 1999–2017.

Benz, Matthias, Frey, Bruno S., 2004. Being independent raises happiness at work. Swedish Economic Policy Review 11 (2), 95–134.

Berglann, Helge, Moen, Espen R., Røed, Knut, Skogstrøm, Jens Henrik, 2011. Entrepreneurship: origins and returns. Labour Economics 18 (2), 180–193.

Black, Sandra, Devereux, Paul, Salvanes, Kjell, 2007. From the cradle to the labor market? The effect of birth weight on adult outcomes. Quarterly Journal of Economics 122 (1), 409–439.

Blanchflower, David, 2000. Self-employment in OECD countries. Labour Economics 7 (5), 471–505.

Blanchflower, David, Oswald, Andrew, 1998. What makes an entrepreneur? Journal of Labor Economics 16 (1), 26–60.

Bonjour, Dorthe, Cherkas, Lynn, Haskel, Jonathan, Hawkes, Denise, Spector, Tim, 2003. Returns to education: evidence from U.K. twins. American Economic Review 93 (5), 1799–1812.

Bound, John, Solon, Gary, 1999. Double trouble: on the value of twins-based estimation of the return to schooling. Economics of Education Review 18 (2), 169–182.

Bruce, Donald, Schuetze, Herbert, 2004. The labour market consequences of experience in self-employment. Labour Economics 11 (5), 575–598.

Carlin, John, Gurrin, Lyle, Sterne, Jonathan, Morley, Ruth, Dwyer, Terry, 2005. Regression models for twin studies: a critical review. International Journal of Epidemiology 34 (5), 1089–1099.

Carrington, William J., McCue, Kristin, Pierce, Brooks, 1996. The role of employer/employee interactions in labour market cycles: evidence from the self-employed. Journal of Labour Economics 14 (4), 571–602.

Chamberlain, Gary, 1980. Analysis of covariance with qualitative data. Review of Economic Studies 47 (1), 225–238.

DeFries, J., Fulker, D., 1985. Multiple regression analysis of twin data. Behavior Genetics 15 (5), 467–473.

Dunn, Thomas, Holtz-Eakin, Douglas, 2000. Financial capital, human capital and the transition to self-employment: evidence from intergenerational links. Journal of Labor Economics 18 (2), 282–305.

Evans, David, Leighton, Linda, 1989. Some empirical aspects of entrepreneurship. American Economic Review 79 (3), 519–535.

Fairlie, Robert W., 1995. Self-employment, entrepreneurship, and the NLSY79. Monthly Labor Review 40–47 (February).

Ferrer-i-Carbonell, A., Frijters, P., 2004. How important is methodology for the estimates of the determinants of happiness? The Economic Journal 114 (497), 641–659.

Griliches, Zvi, 1979. Siblings models and data in economics: beginnings of a survey. Journal of Political Economy 87 (5), S37–S64.

Gurrin, Lyle, Carlin, John, Sterne, Jonathan, Dite, Gillian, Hopper, John, 2006. Using bivariate models to understand between- and within-cluster regression coefficients, with application to twin data. Biometrics 62 (3), 745–751.

Hamilton, Barton, 2000. Does entrepreneurship pay? An empirical analysis of the returns to self-employment. Journal of Political Economy 108 (3), 604–631.

Hundley, Greg, 2001. Why and when are the self-employed more satisfied with their work? Industrial Relations 40 (2), 293–316.

Hyytinen, Ari, Rouvinen, Petri, 2008. The labour market consequences of self-employment spells: European evidence. Labour Economics 15 (2), 246–271.

Isacsson, Gunnar, 2007. Twin data vs. longitudinal data to control for unobserved variables in earnings functions — which are the differences? Oxford Bulletin of Economics and Statistics 69 (3), 339–362.

Johansson, Edvard, 2005. An estimate of self-employment income underreporting in Finland. Nordic Journal of Political Economy 31, 99–110.

Kaprio, Jaakko, Koskenvuo, Markku, 2002. Genetic and environmental factors in complex diseases: the older Finnish twin cohort. Twin Research 5 (5), 358–365.

Kaprio, Jaakko, Koskenvuo, Markku, Artimo, Markus, Seppo, Sarna, Rantasalo, Ilari, 1979. The Finnish Twin Registry: Baseline Characteristics. Section I. Materials Methods, Representativeness and Results for Variables Special to Twin Studies: Kansanterveystieteen julkaisuja M 47.

Kawaguchi, Daiji, 2008. Self-employment rents: evidence from job satisfaction scores. Hitotsubashi Journal of Economics 49 (1), 35–45.

Korenman, Sanders, Neumark, David, 1991. Does marriage really make men more productive? Journal of Human Resources 26 (2), 282–307.

Li, Hongbin, Liu, Pak-Wai, Zhang, Junsen, 2012. Estimating returns to education using twins in urban China. Journal of Development Economics 97 (2), 494–504.

Neumark, David, 1999. Biases in twin estimates of the return to schooling. Economics of Education Review 18 (2), 143–148.

Nicolaou, Nicos, Shane, Scott, 2009. Can genetic factors influence the likelihood of engaging in entrepreneurial activity? Journal of Business Venturing 24 (1), 1–22.

Nicolaou, Nicos, Shane, Scott, Cherkas, Lynn, Hunkin, Janice, Spector, Tim D., 2008a. Is the tendency to engage in self-employment genetic? Management Science 54 (1), 167–179.

Nicolaou, Nicos, Shane, Scott, Cherkas, Lynn, Spector, Tim D., 2008b. The influence of sensation seeking in the heritability of entrepreneurship. Strategic Entrepreneurship Journal 2 (1), 7–21.

Parker, Simon C., 2009. The Economics of Entrepreneurship. Cambridge University Press, Cambridge.

Shane, Shane, Nicolaou, Nicos, Cherkas, Lynn, Spector, Tim, 2010. Genetics, the big five, and the tendency to be self-employed. Journal of Applied Psychology 95 (6), 1154–1162.

Silventoinen, Karri, Sarlio-Lähteenkorva, Sirpa, Koskenvuo, Markku, Lahelma, Eero, Kaprio, Jaakko, 2004. Effect of environmental and genetic factors on education associated disparities in weight and weight gain: a study of Finnish adult twins. American Journal of Clinical Nutrition 80 (4), 815–822.

Taubman, Paul, 1976. The determinants of earnings: genetics, family and other environments; a study of American white male twins. American Economic Review 66, 858–870.

Taylor, Mark, 1996. Earnings, independence or unemployment: why become self-employed. Oxford Bulletin of Economics and Statistics 58 (2), 253–266.

Taylor, Mark, 1999. Survival of the fittest? An analysis of self-employment duration in Britain. The Economic Journal 109 (454), C140–C155.

Taylor, Mark, 2004. Self-employment in Britain: when, who and why? Swedish Economic Policy Review 11 (2), 139–173.

Tergiman, Chloe, 2011. Entrepreneurship does pay. Working Paper. University of British Columbia, Vancouver.

Uusitalo, Roope, 2001. Homo entreprenaurus? Applied Economics 33 (13), 1631–1638.

Van der Loos, Matthijs, Koellinger, Philipp, Groenen, Patrick, Thurik, Roy, 2010. Genome-wide association studies and the genetics of entrepreneurship. European Journal of Epidemiology 25 (1), 1–3.

Van der Loos, Matthijs, et al., 2011. Candidate gene studies and the quest for the entrepreneurial gene. Small Business Economics 37 (3), 269–275.

Wooldridge, Jeffrey M., 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge, Massachusetts.

Zhang, Junsen, Liu, Pak-Wai, Yung, Linda, 2007. The Cultural Revolution and returns to schooling in China: estimates based on twins. Journal of Development Economics 84 (2), 631–639.