Discussion Papers
Statistics Norway
Research department
No. 789
November 2014

Statistics Norway

Statistisk sentralbyrå

Mirjam van Praag and Arvid Raknerud

The entrepreneurial earnings puzzle

Evidence from matched person-firm data

Discussion Papers No. 789, November 2014
Statistics Norway, Research Department

Mirjam van Praag and Arvid Raknerud

The entrepreneurial earnings puzzle
Evidence from matched person-firm data

Abstract: Empirical studies show that the pecuniary returns to an individual's decision to switch from wage employment to entrepreneurship are low. We reconsider the pecuniary gains from this transition using a unified and flexible approach based on a mixed model with heterogeneous returns to entrepreneurship. Addressing the issue of self-selection, we analyze to what extent earlier findings are obscured by mixing individuals who become entrepreneurs without interesting wage alternatives with those who do have a realistic alternative opportunity. Our data set covers the whole Norwegian population of individuals matched to the entire population of firms established in the period 2002--2011, and includes extensive income and ownership share measures. The results indicate that the average return to entrepreneurship is significantly negative for individuals entering entrepreneurship through self-employment. Entrepreneurs who establish firms by injecting the minimum (or close to minimum) required amount of equity in an incorporated firm at start-up, have a significantly positive, but low return to entrepreneurship on average. Finally, persons who become entrepreneurs by establishing firms that are at least twice as large as the minimum requirement, increase their earnings by 10 percent on average by becoming entrepreneurs. We identify a significant positive selection by absolute advantage with regard to the choice of becoming an incorporated entrepreneur, but not with regard to self-employment.

Keywords: Entrepreneurship, Returns to entrepreneurship, Earnings distribution, Matched person-firm data

JEL classification: L26, C23, J31

Acknowledgements: We appreciate useful comments from Torbjørn Hægeland and Terje Skjerpen, and invaluable research assistance from Diana Iancu. Financial support from the Norwegian Research Council through the program MER Entreprenørskap, grant no 201340, is gratefully acknowledged.

Address: Mirjam van Praag, Copenhagen Business School, Dept. of Innovation and Organizational Economics. E-mail: mvp.ino@cbs.dk

Arvid Raknerud, Statistics Norway, Research Department. E-mail: rak@ssb.no

Discussion Papers	comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no/en/forskning/discussion-papers
http://ideas.repec.org/s/ssb/dispap.html

ISSN 1892-753X (electronic)

Sammendrag

Tidligere studier viser at pekuniær belønning generelt ikke er hovedmotivasjonen for overganger fra vanlig lønnsarbeid til entreprenørskap. Vi revurderer i denne analysen de pekuniære gevinstene og risikoen ved entreprenørskap, ved å bruke et enhetlig og fleksibelt økonometrisk rammeverk. Vi tar hensyn til selv-seleksjon og undersøker i hvilken grad tidligere analyser er kommet til feilslutninger fordi en har blandet sammen individer uten interessante lønnsalternativer med de som har en høy alternativ inntekt i arbeidsmarkedet når beslutningen om etablering tas. Vi analyserer registerdata som omfatter hele den norske sysselsatte befolkingen i perioden 2001-2011 og definerer en entreprenør som en person som enten er innehaver av et nyetablert enkeltmannsforetak (selvstendig næringsdrivende), eller er aktiv eier i et nyetablert aksjeselskap (AS) – i det siste tilfellet har han/hun minst blokkerende mindretall (33 prosent eierandel) og er i tillegg enten ansatt i foretaket eller har en formell rolle som daglig leder, styreformann eller begge. Våre resultater viser at den gjennomsnittlige «avkastningen» av entreprenørskap er signifikant negativ blant de som blir selvstendig næringsdrivende (etablerer enkeltmannsforetak). Videre finner vi at entreprenører som etablerer foretak ved å skyte inn minimumskapitalkravet (eller nær minimumskravet) i et AS har en signifikant positiv, men beskjeden, avkastning av etableringen. Til slutt finner vi at gruppen entreprenører som skyter inn minst to ganger minimumskravet i et nystartet AS, i gjennomsnitt øker sin (arbeidsrelaterte) inntekt med 10 prosent sammenlignet med hva den ville ha vært dersom de hadde fortsatt som vanlige lønnsmottakere.

1 Introduction

A large literature has emerged examining returns to entrepreneurship (Aastebro, 2010; Aastebro and Chen, 2014; Berkhout et al., 2011). A seminal contribution is Hamilton (2000), who examines differences in the earnings distribution of wage-earners and self-employed persons using a traditional Mincer-type earnings equation framework. He finds that entrepreneurs have, cet. par., lower initial earnings and lower earnings growth than wage-earners and that their earnings distribution exhibits wider dispersion. Hamilton also shows that the earnings differential is not due to self-selection of low ability employees into entrepreneurship. The consensus based on Hamilton's study and follow-up studies is that the pecuniary returns to entrepreneurship are not the driving force of an individual's decision to switch from wage employment to entrepreneurship (Astebro, 2010: Astebro and Chen, 2014). Four sets of discussions to better understand this "entrepreneurial earnings puzzle" have developed since.

The first is that non-pecuniary benefits from entrepreneurship must be substantial (e.g., Benz and Frey, 2008; Blanchflower and Oswald, 1998; Carter, 2011; McCraffrey, 2014). Second, other factors than mere rational expectations might lead to the choice of entrepreneurship such as genetic and environmental factors (Lindquist et al., 2015; Nicolaou et al., 2008) or cognitive biases (Holm et al., 2013) arising from, for instance, overoptimism (Lowe and Ziedonis, 2006; Dushnitsky, 2010) and/or overconfidence (Hayward et al., 2006) or a lower level of risk or loss-aversion (Hvide and Panos, 2014; Koudstaal et al., 2014).

Third, researchers have also attributed the lack of evidence of an effect of entrepreneurial income to the low quality of entrepreneurial income data (Parker, 2009). Indeed, measuring business incomes is notoriously difficult due to a lack of unequivocal accounting and reporting methods and misreporting (Astebro and Chen, 2014; Astebro, 2010; Feldman and Slemrod, 2007; Hurst et al., 2013). Moreover, the range of possible outcomes is much wider and often not even foreseeable due to risk and uncertainty (Astebro, 2010, Parker, 2009, Hamilton, 2000). No doubt, the lack of a uniform concept of entrepreneurial income is also a problem for individuals who consider starting up a venture. Supporting this idea, Berkhout et al. (2011) show for a Dutch sample of young college and university graduates that decisions to become entrepreneurs are indeed not affected by the entrepreneurial income prospects. However, the decision to become an entrepreneur is affected significantly by the income prospects in wage-employment in their labor market segment, i.e., the better observable and measurable opportunity costs of entrepreneurship.

Fourth, a new perspective is provided based on the debate about 'who is an entrepreneur?'. This perspective suggests the use of 'stricter' definitions of entrepreneurship, thereby weeding out from the sample "necessity" entrepreneurs and those self-employed who typically earn low incomes and experience little growth. The resulting population of entrepreneurs would be more representative of the population of true 'Schumpeterian' entrepreneurs. Definitions that have been used for this purpose are, for instance, incorporated entrepreneurs (Levine and Rubinstein,

2013), offensive entrepreneurs, i.e., movements from wage-employment to entrepreneurship (thereby ignoring entrepreneurship originating from unemployment; see Berglann et al., 2009), or even billionaire entrepreneurs (Henrekson and Sanandaji, 2014). Levine and Rubinstein show that incorporated business owners earn, on average, more than both unincorporated (self-employed) business owners and ordinary wage workers. Wage workers earn, in turn, more than self-employed workers.

The main aim of our analysis is to contribute to a better understanding of the 'entrepreneurial earnings puzzle' than currently available in the literature. We try to do so based on two pillars: (i) data improvements and (ii) model improvements. The data improvements are that we have a large and relatively long panel data set covering the whole Norwegian population of firms and individuals from 2001 to 2011, where we can observe the initial earnings for given individuals before they become entrepreneurs. In our analysis the initial condition is that the person is a wage employer in 2001. Hence, we can focus on offensive entrepreneurship. In contrast, self-employment as an alternative to unemployment or social benefits is likely to be less influenced by future earnings prospects, and should be considered separately (Hvide and Panos, 2014). Second, we observe very detailed measures of both employment and entrepreneurship incomes in the registry. For the latter we have detailed information on share ownership and valuation leading to the possibility of measuring firm value growth and income from ownership in the case of incorporation (see also Hvide and Møen, 2010, using similar Norwegian registry data). We acknowledge the self-selection problem and analyze to what extent earlier findings are obscured by mixing individuals who become entrepreneurs without interesting wage alternatives (low ability) with those who do have a realistic alternative opportunity (high ability). We distinguish between incorporated and unincorporated (self-employed) entrepreneurs.

Our estimation model has the following features: (i) It can be used to identify whether persons who have a high general income ability are the ones who choose to become entrepreneurs ("selection by absolute advantage"). We measure income ability as the part of the income which is unchanged over time, irrespective of the choice to become an entrepreneur. (ii) We can assess the average *return* to entrepreneurship for those who become entrepreneurs, i.e. the increase in their earnings by becoming entrepreneurs, compared to remaining wage earners ("the average treatment effects on the treated"). (iii) It allows entrepreneurial and wage incomes to differ not only in terms of their expected value but also in terms of variance. (iv) The choice to become an entrepreneur is considered endogenous with respect to income prospects.

Using this combination of data and model improvements we show that the return to entrepreneurship is indeed positive when using a stricter definition of the entrepreneur. While the average return to entrepreneurship is significantly negative for individuals entering entrepreneurship through self-employment (even if some of them later incorporate their firm), entrepreneurs who establish firms by injecting the minimum (or close to minimum) required amount of equity in an incorporated firm at start-up, have a significantly positive, but low return to entrepreneurship on average. However, people who become entrepreneurs by establishing firms that are at least twice as large as the minimum requirement, increase their earnings by 10 percent on average by becoming entrepreneurs. Besides, we identify a signifi-

cant positive selection by absolute advantage with regard to the choice of becoming an incorporated entrepreneur. The estimated returns to entrepreneurship are not significantly different for men and women.

The remainder of the paper is organized as follows. In Section 2 we discuss the modelling framework and in Section 3 the data. In Section 4 we discuss the results. Section 5 concludes.

2 The modelling framework

In this section we specify the modelling framework for estimating the earnings equation and the choice equation of whether to become an entrepreneur. In principle, we could use our panel data set to study transitions both into and out of entrepreneurship over time. However, to simplify the analysis, we will focus on one type of transition: from initial full-time employment in 2001, defined as 30 hours or more per week, to entrepreneurship in any of the years 2002-2011. The initial condition is that the individual is a full time wage earner, thereby excluding unemployed individuals from the sample. The motivation for this choice is that we wish to concentrate on offensive entrepreneurship choices rather than necessity or defensive entrepreneurship. Otherwise, it is difficult to study the relative returns to the choice for entrepreneurship.

The discrete choice we consider is the decision to become an entrepreneur during the observation period, which is 2002-2011, given the initial condition in year 2001. Thus the model assumes that the decision to become an entrepreneur is static, and is related only to the initial or exogenous characteristics of the individual. The actual timing of events (the *years* of transitions from one state two another) will be considered as random conditional on this discrete choice. Some will become entrepreneurs late in the period, and we may follow them as entrepreneurs only for a year or two. Those who make the transition early can potentially be observed for a longer time – but some of them will exit entrepreneurship before the period is over. Of course, a large majority of wage earners never become entrepreneurs. When a transition occurs, there will be a shift in the earnings equation. The endogeneity of the decision to become an entrepreneur means that pre- and post- decision earnings may be correlated with the decision to become an entrepreneur.

2.1 Stochastic specification

Individual i's state at time t is denoted $Ent(i,t) \in \{0,1\}$, where $Ent(i,t) = 1$ means that the individual is an entrepreneur at time t and $Ent(i,t) = 0$ means that he is not. For now we do not separate between different types of entrepreneurship, but will return to this extension of the model at the end of the section. Initially $Ent(i,0) = 0$, for all individuals in the sample, i.e., they are full time wage earners at $t = 0$.

The discrete choice variable, E_i, is defined as

$$E_i = \max_{t \in [1,T]} Ent(i,t).$$

A typical event history may be i) $Ent(i,t) = 0$ for all t (the person remains a wage earner), ii) $E(i,t)$ jumps from 0 to 1 at t' (entry into entrepreneurship), and iii) $E(i,t)$ drops back from 1 to 0 at $t'' > t'$ (exit from entrepreneurship, see Section 3 for its exact operationalization). To model the discrete choice variable E_i we follow the framework of Dagsvik et al. (2011), which incorporates also an *ordinal* choice variable E_i (see below): Let X_i^* be a continuous latent index representing both individual i's preferences and his opportunities of becoming an entrepreneur. The endogenous choice variable E_i is related to X_i^* by

$$E_i = e \text{ iff } \mu_e < X_i^* < \mu_{e+1} \text{ , } e = 0, 1, \tag{1}$$

where μ_1 is an unknown threshold value, $\mu_0 = -\infty$ and $\mu_2 = \infty$. Furthermore, we assume that

$$X_i^* = Z_{1i}\gamma_1 + \varepsilon_{1i}, \tag{2}$$

where Z_{1i} is a row vector of exogenous variables affecting the individual's choice (for instance age and gender) pre-determined at $t = 0$, γ_1 is a fixed, unknown coefficient vector and ε_{1i} is a normally distributed random variable with zero mean and unit variance. It is assumed that ε_{1i} is uncorrelated with Z_{1i} by construction. That is, (2) is the conditional distribution of X_i^* given Z_{1i}. We are here not interested in the causal effect of Z_{1i} (including e.g. education) on the choice to become an entrepreneur per se. However, we will allow ε_{1i} to be correlated with earnings. That is, any unobserved variable that affects the entrepreneurship decision is allowed to influence earnings. This creates a self-selection problem that will be addressed below. Equations (1)-(2) specify a Probit model for the binary choice variable, E_i. The estimation of such a model is, of course, standard.

Our equation of main interest is the earnings relation. Our main focus will be on the returns to entrepreneurship, i.e., the change in total labor related earnings (wage-, entrepreneurial- and other) when a person switches from full time wage employment to entrepreneurship during a time period. We propose a model which is analogous to the two-sector model of Heckman and Sedlacek (1990). Let the index i denote individual i, and $Y_{it}(0)$ and $Y_t(1)$ log earnings if individual i at t when $Ent(i,t) = 0$ and $Ent(i,t) = 1$, respectively. Hence, log earnings for a person who becomes an entrepreneur at t, switches from $Y_{i,t-1}(0)$ at $t-1$ to $Y_{i,t}(1)$ at t. We assume that

$$\begin{aligned} Y_{it}(0) &= Z_{2it}\gamma_2 + \varepsilon_{2i} + u_{it}(0) \\ Y_{it}(1) &= \beta_i + Z_{2it}\gamma_2 + \varepsilon_{2i} + u_{it}(1), \end{aligned} \tag{3}$$

where β_i is the shift in income when person i changes from being a wage earner to an entrepreneur. Z_{2it} is a vector of explanatory variables, including (powers) of years of experience, calendar time dummies, as well as variables from Z_{i1}. Experience is measured as age minus years of schooling (minus seven years), and thus reflects *potential* experience. The restriction that the coefficients of Z_{2it} (i.e., γ_2) are the same in both equations facilitates the interpretation of the results. Finally, ε_{2i} is a person-specific random effect and $u_{it}(0)$ and $u_{it}(1)$ are the idiosyncratic error terms in state $Ent(i,t) = 0$ and $Ent(i,t) = 1$, respectively.

To allow heterogeneity, β_i is an individual-specific coefficient, with

$$E(\beta_i | E_{nt}(i,t) = 1) = E(\beta_i | E_i = 1) = \beta. \tag{4}$$

That is, β is the average treatment effect on the treated.

The (hypothetical) earnings difference for the same person i in state 1 and 0, i.e., his returns to entrepreneurship (Δ_i), equals

$$\Delta_{it} = \beta_i + u_{it}(1) - u_{it}(0).$$

Thus the average earnings difference (average treatment effect) is

$$AT \equiv E(\beta_i) \tag{5}$$

and the average treatment effect given treatment is

$$ATT \equiv E(\Delta_{it}|E = 1) = E(\beta_i|E_i = 1) = \beta \tag{6}$$

The average observed earnings differentials between entrepreneurs and non-entrepreneurs (OD) – everything else equal – is given by

$$OD \equiv ATT + E(\varepsilon_{2i}|E_i = 1) - E(\varepsilon_{2i}|E_i = 0). \tag{7}$$

ATT is higher than AT if the idiosyncratic component in the returns to entrepreneurship (β_i) is on average higher for those who actually choose to become entrepreneurs than for the average individual. We characterize this as selection by comparative advantage. Equation (7) shows that the correlation between the additive individual effect in the earnings equation (ε_{2i}) and the error term in the choice equation (ε_{1i}) prevents us from estimating ATT simply from observed earnings differentials between entrepreneurs and non-entrepreneurs (even if we control for differences in observed variables). The bias may be positive or negative: Those with a higher earnings potential regardless of entrepreneurship may tend to become entrepreneurs (such positive correlation between ε_{2i} and ε_{1i} can be interpreted as selection by *absolute advantage*), or they may tend to remain wage earners (negative correlation). Thus it is adamant to be able to control for self-selection when making inferences about treatment effects.

To address selection effects discussed above, we allow ε_{1i}, ε_{2i} and β_i to be correlated random variables. However, since we will not attempt to estimate AT – only ATT – we do not explicitly model the correlation between ε_{1i} and β_i (more about this below). To estimate equation (3), we will condition on the choice variable E_i, and obtain expressions for the conditional expectations of ε_{2i} given E_i. To do so, we first note that we can write

$$\varepsilon_{2i} = \theta_1 \varepsilon_{1i} + \widetilde{\varepsilon}_{2i}, \tag{8}$$

where $\widetilde{\varepsilon}_{2i}$ and ε_{1i} are independent and

$$E(\varepsilon_{1i}\varepsilon_{2i}) = \theta_1. \tag{9}$$

Defining $Y_{it}(E(i,t)) \equiv Y_{it}$ and $u_{it} \equiv u_{it}(Ent(i,t))$, the observed time series is

$$Y_{it} = \beta_i Ent(i,t) + Z_{2it}\gamma_2 + \varepsilon_{2i} + u_{it}, . \tag{10}$$

Only one of the potential outcomes $Y_{it}(0)$ and $Y_{it}(1)$ is observed – the other is a counterfactual outcome. To estimate (10) given that $Ent(i,t)$ is endogenous and

depends on ε_{2i}, we apply a control function approach, in the tradition of Heckman (1979) and Garen (1984). That is, we will derive auxiliary variables that are computed from the Probit analysis described above and include these as an additional regressor (control function) in the earnings equation (3) thereby accounting for the correlation between $Ent(i,t)$ and ε_{2i}. The original earnings equation can then be transformed into an equation with a genuine random effect that is uncorrelated to the explanatory variables. Our main result to this effect is in Proposition 1.

Proposition 1 *Assume that* $(\varepsilon_{1i}, \varepsilon_{2i})$ *is binormally distributed with zero mean and satisfies the conditions specified above, given by (8) and assume that E is determined by the probit model (1)-(2). Then*

$$E(\varepsilon_{2i}|E_i = e) = \theta_1 \xi_i(e)$$

where

$$\xi_i(e) = \frac{\phi\left(\mu_{e-1} - Z_{1i}\gamma_1\right) - \phi\left(\mu_e - Z_{1i}\gamma_1\right)}{\Phi\left(\mu_e - Z_{1i}\gamma_1\right) - \Phi\left(\mu_{e-1} - Z_{1i}\gamma_1\right)}.$$

The proof is standard, but stated in the Appendix for completeness. It follows that we can express (10) as

$$Y_{it} = \beta Ent(i,t) + \delta_i Ent(i,t) + Z_{2it}\gamma_2 + \theta_1 \xi_i(e) + \varepsilon_i^* + u_{it}, \tag{11}$$

where

$$\begin{aligned} \delta_i &= \beta_i - \beta \\ \varepsilon_i^* &= \varepsilon_{2i} - \theta_1 \xi_i(e) \end{aligned}$$

and the random effect, ε_i^*, has the property that $E(\varepsilon_i^*|E_i = e) = 0$. Moreover, we allow heteroscedasticity in the random effect, ε_i^*, and an autoregressive structure in the error term, u_{it} :

$$\begin{aligned} &Var(\varepsilon_i^*|E_i = e) = \sigma^2(e),\ e = 0, 1 \\ &u_{it} = \phi u_{i,t-1} + \eta_{it}, \eta_{it} \sim i.i.d(0, \sigma_\varepsilon^2) \end{aligned}$$

Equation (11) is a *mixed* model, with β $(= ATT)$ as the fixed interest parameters and δ_i as the random parameters. To estimate β consistently by standard mixed models methods, the following assumptions must hold:

Assumption 1 $E(\delta_i u_{it}) = 0$ for all t

Assumption 2 $E(\delta_i\ Ent(i,t)) = 0$ for all t

Assumption 3 $E(u_{it} Ent(i,t)) = 0$ for all t

Assumption 1 is standard, the firm-specific effect of the treatment (measured as a deviation from the average treatment effect on the treated) must not depend on any of the genuine error terms. Assumption 2 is satisfied given our previous assumptions: Given (4), $E((\beta_i - \beta)|Ent(i,t) = 1) = E(\delta_i|Ent(i,t) = 1) = 0$. Moreover,

$$E(\delta_i|Ent(i,t) = 1) = 0 \Rightarrow E(\delta_i Ent(i,t)) = 0 \tag{12}$$

Assumption 3 comes down to independence of the treatment from the genuine error term. The selection effects are captured by the control function $\xi_i(e)$, entering as a linear regressor into the model. Note however, that the model allows δ_i to be correlated with ε_i^*.

Equation (11) is a standard mixed model, with a random coefficient corresponding to the variable $Ent(i,t)$, a random intercept ε_i^*, and an autoregressive error term, u_{it}. A positive θ_1 (positive correlation between ε_{1i} and ε_{2i}) can be interpreted as "selection by absolute advantage", whereas a positive correlation between ε_{1i} and β_i can be interpreted as "selection by comparative advantage".

It follows from (7) that

$$OD \equiv ATT + \theta_1 E(\xi_i(1) - \xi_i(0)). \tag{13}$$

Thus, to calculate treatment effects, the parameters β and θ_1 and the control function $\xi_i(e)$ are of interest.

A key issue is the search for valid exclusion restrictions (exogenous variation in the discrete choice variable) with the purpose of identifying key structural parameters associated with the returns to entrepreneurship.[1] We discuss and implement one exclusion restriction in the equation for the entrepreneurship choice (E): whether a non-parent family member (e.g. brother or sister) is an entrepreneur (Lindquist et al., 2015). Family entrepreneurship increases the probability of becoming an entrepreneur, but does not affect entrepreneurial earnings as long as parental entrepreneurship is excluded from the set of identifying instruments (and showing up in both equations). Parental entrepreneurship is associated with the transmission of human and financial capital that possibly may have a direct effect on earnings.

The regression parameters in the above model can be estimated by standard random effects methods, such as quasi-maximum likelihood estimation based on the assumption that all random terms are normally distributed. The reason is that the first order conditions are correctly specified and that the normal distribution is a member of the exponential family (see Gourieroux and Montfort, 1995).[2]

The above framework estimates an average treatment effect β (relative shift in income) for all entrepreneurs irrespective of their initial investment in the firm.[3] It is quite easy within our current framework to allow heterogeneity in the treatment effects by letting ATT to depend on the initial investment. For example, define

$$E_i = n \text{ iff } \mu_n < X_i^* < \mu_{n+1} \text{ , } n = 0, 1, ..., K, \tag{14}$$

where $E_i = n$, for $n \in \{1, ..., K\}$, means that the individual is an entrepreneur and that the initial injected equity in the firm[4] lies in some interval $(K_{n-1}, K_n]$. As above, $E_i = 0$ for non-entrepreneurs. Define corresponding variables $Ent^{(n)}(i,t)$, $n = 1, ..., K$, which take the value 1 in all years when the individual is an entrepreneur, given that his *initial* investment is n (ordinal variable), and 0 else. The average

[1] Within our framework such restrictions are not formally needed to obtain identification.

[2] We estimate the model using the *mixed* command in STATA.

[3] Even though we subtract a normal return on equity in our earnings measure, it may be the case that individuals who start up large firms are systematically different (e.g. require a higher risk premium) than those who start up smaller firms, or invest less.

[4] The minimum required injected equity for a firm is NOK 100,000.

treatment effect on the treated corresponding to $E_i = n$ can then be written as

$$ATT(n) \equiv \beta^{(n)} = E(\beta_i^{(n)} | E_i = n).$$

We will present results for a version of this generalized (ordinal) model with $K = 2$, and where $E_i = 1$ means that the injected initial equity is less than NOK 200,000 and $E_i = 2$ that the injected initial equity is at least 200,000. The highest category comprises 1/3 of all incorporated entrepreneurs.

3 Definitions and data

The entrepreneur Most empirical studies have measured entrepreneurship in terms of self-employment. This has lately become a much debated choice (Levine and Rubinstein, 2013; Henrekson and Sanandaji, 2014). A concept of entrepreneurship that does not include incorporated firms will miss out on the most successful entrepreneurs. Official statistics confirm that self-employment has not been an important source of labor income growth in Norway during the last decades, in contrast to wage- and business income from incorporated firms.[5]

Using Norwegian registry data, Berglann et al. (2011) invoke a wider definition of entrepreneurship. An entrepreneur is either employed in a firm in which (s)he is a major/active owner (with at least 30 percent ownership or a combination of +10 percent ownership and being a board member or a chief executive) or who runs his or her own business as a sole proprietor. In other studies using Norwegian registry data, Hvide (2009) and Hvide and Panos (2014) define an entrepreneur as an individual with a majority stake, i.e., more than 50% of the total shares, in a newly established incorporated company. Both these definitions have limitations. Regarding the owner-based definition, Luger and Koo (2005) point out that many new firms are simply continuations of existing firms through a judicial reorganization. Moreover, a large share of new firms are merely formal entities without real economic activity (although Hvide, 2009, controls for this by imposing a minimum threshold on the activity level).

We distinguish two types of 'entrepreneurship': self-employment and incorporation.[6] Similar to Berglann et al. (2011), we define an entrepreneur based on a combination of ownership and control in a company. The entrepreneur must have at least a blocking minority position in a privately held limited liability company (>33%) and, at the same time, must be either an employee or have a formal management role (CEO, chairman of the board, or both) during start-up. The 33% threshold includes both direct and indirect ownership positions in the firm (more about this below)[7]. The choice of a threshold necessarily involves some arbitrariness, but our criteria ensure that the entrepreneur retains a certain degree of control over the firm and, at the same time, is an active owner. For entrepreneurs in the sense of

[5] See Fjærli et al., 2013, Figure 1.

[6] The latter category will later be divided into two start-up size groups.

[7] We acknowledge that the decision to start an incorporated firm is sometimes motivated by tax planning, rather than entrepreneurship. At the same time, self- employment may be a close substitute for wage employment, typically in the low end of the wage distribution and may also have little to do with entrepreneurship in the classical, Schumpeterian sense.

self-employment, we require that they run their own business as a sole proprietor. For both types of entrepreneurship, we require that the firm is new. That is, persons who become owner-managers or sole proprietors of already existing firms (e.g. who take over a family business) are not classified as entrepreneurs.[8]

The earnings measure In contrast to a passive portfolio investor and ordinary wage earner, the incorporated entrepreneur invests both financial and human capital in one and the same firm. In most cases, the owner–manager will have full control and can easily transfer equity in and out of the firm without regarding the preferences of other shareholders and conflicts of interest. This also means that the entrepreneur can decrease (increase) the level of profit in the firm by increasing (decreasing) her own wage, possibly motivated by tax concerns (Astebro and Chen, 2014; Astebro, 2010; Feldman and Slemrod, 2007; Hurst et al., 2013).

Total earnings as the sum of labor income and ownership income from the entrepreneur's own firm are of interest when estimating the pecuniary returns to entrepreneurship. Unfortunately, realized business income (such as dividends) from corporate firms is not a good measure of entrepreneurial earnings. First, they are vulnerable to changes in taxation rules – of particular importance for Norway in the last decade where the pre-announced tax reform of 2006, for instance, led to a huge step-up of dividend payments in 2005, and a sharp decline in the subsequent years.[9] An additional problem is that business income for an entrepreneur will typically consist of both returns to invested financial capital (equity) and returns to human capital (labor and effort).

We propose a uniform measure of (pretax) earnings for all individuals in the sample, whether self-employed, wage-employed, unemployed or owner-managers of incorporated firms. It is the sum of labor income (from wages and self-employment), work-related cash transfers (such as unemployment benefits and short-term sickness benefits) and owner income from incorporated firms in which the individual is an entrepreneur. The latter is denoted "entrepreneurial owner income" (in contrast to general capital income, which is not included in our earnings measure, because it is a return on a portfolio investment, not entrepreneurial effort). Entrepreneurial owner income is allocated to the owners of a firm in proportion to their ownership share, and is defined as total taxable profit in a specific year *after subtracting a normal rate of return to the firm's equity* (injected equity plus accumulated retained earnings). The latter is done to account for the opportunity cost of invested financial capital, which should not be counted as a part of the return to entrepreneurship (cf. Parker, 2009). The normal return is simply set to 4% which was the average (nominal) yield on 10 year government bonds during 2001-2011.

Our approach is in the spirit of Hamilton (2000), but is based on much more detailed accounting and ownership data at the firm and person level. It will, nevertheless, still have weaknesses, for example because the firm may possess hidden values not yet materialized as profits that may not be revealed until shares are sold.

[8] An entrepreneur may still be a wage earner in another firm than where (s)he is an owner-manager.

[9] Also in 2002, when an increase in dividend taxation was largely expected, there was a bust in dividend payments in an otherwise poor year for business owners.

Identification of owners and ownership shares in the registry We focus on private firms registered under the organizational forms AS (aksjeselskap/private limited company/incorporated firms) and ENK (sole proprietorship/self-employment) between 2001 and 2011.[10] We use data from different registers, covering the entire population of firms and owners. These are:

- The Household register[11]. This register includes a wealth of information about individuals and households obtained by merging several primary registers. It contains annual information about income, wealth, education, and demographic variables, including identification numbers of individuals' spouses and relatives, for all persons above the age of 18 with permanent residency in Norway.

- The Directorship register. This provides details for each individual appointment in positions such as general manager, chairman or member of the board for AS firms and sole proprietors in the case of ENK firms.

- The Register of Employers and Employees. This contains data on employment contract durations, wage and contractual working hours for each employee, including sole proprietors (ENK firms).

- The Shareholder register. This register provides information about owners (both individuals and firms) and their shareholdings from 2001 and onwards.[12]

- The Accounts statistics. This register contains data from the financial statements of AS firms.

- The Central register of establishments and enterprises, with information about the establishment and termination of all registered firms (date of establishment, date of closure, reason for closure (e.g. bankruptcy, merger, overtaken by another firm, or unspecified). The register also includes information on the firm's industry (4-digit NACE), number of employees, turnover (total sales) and location.

To identify entrepreneurs in the sense of sole proprietors we use the Directorship register to identify the sole proprietor and the Central register of establishments and enterprises to match the individual sole proprietor to a new firm. The register of establishments and enterprises includes a binary activity code (active or non-active) assigned by Statistics Norway; an active firm is required to have registered some form of economic activity, such as positive turnover (total sales income) or payments of value added tax. Newly established firms without registered activity are not classified as established before they become active. Inactive firms are removed from the analysis. This applies both to AS and ENK-firms, but is especially important to ENK-firms (of which there is a higher share of inactive firms).

[10] AS owners are obliged to inject a minimum capital of 100 000 NOK at start-up and they have no personal liability for the company's obligations.

[11] This is not formally a register, but a data base obtained by linking register based information from several sources, including income and wealth information from tax records.

[12] Measures are slightly different for the 2001-2003 period due to a shift in data source.

Identifying incorporated entrepreneurs is more complicated. We need to identify the owners of newly established firms and their ownership shares. Moreover, we need to determine whether they are employed in the firm or have an appointment in the firm as general manager, chairman or member of the board. An owners ownership shares in a company include both direct and indirect ownership through other firms (see Fjærli et al., 2013, for details).[13]

The procedure applied to identify ultimate owners enables us to differentiate between three levels of ownership. Level 1 represents direct ownership (the individual shareholder owns part of the firm directly), while levels 2 and 3 indicate indirect ownership (with, respectively, one and two firms acting as intermediaries between the ultimate owner and the firm). For about 80% of the firms during 2001–2003 and 90% during 2004–2011, we identify all shareholders, indicating that most firms in Norway are owned directly or indirectly through only one or two intermediary firms.[14]

After identifying the ultimate personal owners and their ownership shares, we merge the resulting databases with the Accounting statistics for the corresponding years to add a series of firm characteristics. In the resulting matched owner-firm data set, we keep only the firms for which accounting information exists. We also exclude owners that cannot be matched with the Households register.

Sample selection Table 1 shows the total number of employed individuals (in 2001) who established incorporated (AS) or unincorporated (ENK) firms (first and second pair of columns, respectively) in the subsequent period 2002-2011.[15] The last pair of columns of the table displays total number of employed individuals in 2001 for men and women, respectively, i.e., individuals with a registered employment relationship according to the Register of employers and employees (including self-employed individuals). Hence the third pair of columns includes the first two. The entrepreneurs that can be identified in the data are displayed in the first row of Table 1: There are 38,225 and 7,561 men and women, respectively, who become founders of new incorporated firms during 2001-2011. The second pair of columns show that there 83,961 men and 38,961 women who become sole properietors of new firms during 2002-2011. Of the incorporated entrepreneurs in the whole population, 3,867 are first registered as self-employed, and therefore included among the self-employed entrepreneurs in Table 1.

[13] We identify several cases of cross-ownership, when firms simultaneously hold shares in each other. However, as there are few such cases (between 308 cases in 2004 and 429 in 2005), and it is difficult to accurately establish who the ultimate owner is, we exclude cross-ownership cases from our study.

[14] The difference between the two periods is due to the change in data sources, which enabled a more accurate identification of ultimate owners after 2003. Unidentified ownership shares may correspond to a foreign or institutional investors. Foreigners cannot be identified through a Norwegian personal number, while institutions (such as enterprises in the public sector) or listed (ASA) firms are not included in our database.

[15] Holding companies (i.e., companies with an ownership share of minimally 90% in at least one other firm in their first or second year of activity) are excluded from the sample. We also exclude firms in Financial intermediation (NACE 10). These mostly have portfolio investments as their main activity. In addition, we exclude firms with an unspecified industry code and firms in the primary industries (Agriculture and Fishing), as is common in analyses of entrepreneurship.

Table 1: **Sample selection: Entrepreneurs and reference population**

	Incorporated entrepreneurs		Self-employed entrepreneurs		Reference population	
	Men	Women	Men	Women	Men	Women
Initial sample size	38,225	7,561	83,961	38,366	1,159,838	1,039,098
Persons excluded because they are:						
- already entrepreneurs	17,456	1,675	23,627	12,786	306,077	122,946
- above 62 years, or recipients of disbaility or retirement pensions	547	341	2,727	1,961	33,721	97,293
- part-time employees	4,763	2,071	23,799	13,124	275,157	401,076
Final sample size	15,459	3,474	33,808	10,495	544,883	417,783

Some of the individuals in the employee-population are owners of incorporated firms or self-employed in 2001. For reasons discussed above, we exclude individuals who establish a firm, but are already entrepreneurs (e.g. serial entrepreneurs, or self-employed entrepreneurs who incorporate their firm). That is, we exclude individuals who are either i) registered as sole proprietors of existing ENK firms in 2001, or ii) have an ownership share exceeding 33% in an existing incorporated firm, see the second row of Table 1. In the third row, we also exclude individuals older than 62 years in 2001 and recipients of disability- or retirement pensions. Finally, we want to employ the initial condition regarding full-time wage employment in 2001. To operationalize the last requirement, we exclude persons who worked less than 30 hours per week on average as a wage employee in 2001. The numbers pertaining to this last exclusion operation are found in the third row of the table. The numbers of valid entrepreneurs in the sample according to our definitions are shown in the bottom row of Table 1; columns 1–4. The final sample includes 15,459 male and 3,474 female incorporated entrepreneurs (direct transitions from employment to incorporated entrepreneurship) and 33,808 male and 10,495 female self-employed entrepreneurs. The last pair of columns in Table 1 shows that the population of individuals who potentially could make the transition from wage-employment consist of 544,883 men and 417,783 women and is referred to as the "Reference population". These are the individuals satisfying all criteria regarding employment status, age, and of not already being an entrepreneur in 2001 (as we have defined it). This sample forms the basis for estimating Probit models explaining the transition from wage-employment to entrepreneurship.

A few additional remarks about the sample construction are in order. When estimating the earnings equation we do not condition on a person being either a full time employee or an active entrepreneur during the whole observation period (i.e. we do not condition on any *future* labor market outcomes when selecting the sample). If a person chooses to work fewer hours, or becomes voluntarily or involuntarily unemployed he remains in the sample (unemployment insurance is included in the labor earnings). However, to avoid complicating issues related to retirement decisions, we censor all earnings observations above the age of 65 years.[16] Moreover, we censor all observations (technically consider them as "missing") in

[16] As shown by Berglann et al., 2011, entrepreneurs' retirement decisions are markedly different from the reference population.

each year that a person obtains social benefits, disability- or retirement pensions, or has earnings below/above a min/max threshold. The thresholds are chosen so as to censor extraordinarily high and negative earnings observations symmetrically: We remove the 0.5 percent lowest and highest observations in the total sample, which amount to a lower and upper thresholds of NOK 10,000 and NOK 2,500,000, respectively.

Propensity Score Matching (PSM) Before we will estimate the earnings equation, we match the two types of entrepreneurs (the two treatment groups), with corresponding control groups of wage earners by means of propensity score matching. Propensity score matching will ensure that the distribution of the vector of (observed) matching variables, S_i, is the same in both the treatment and control group, and that S_i will be independent of the treatment indicator, E_i, in the *matched* sample of treated and controls, as shown by Rosenbaum and Rubin (1983). Our main motivation for matching is to reduce the sample size when estimating the model described in Section 2, which is necessary for computational feasibility. This does not significantly affect the precision of the estimated treatment effects. An important additional advantage is that matching may alleviate biases if the functional form of S_i in the earnings equation is misspecified – for example if the earnings equation is linear in schooling, but (in reality) the marginal returns to schooling are decreasing. Using PSM, the estimate of the treatment effects would still be unbiased in that case because years of schooling is independent of E_i in the *matched* sample. In the unmatched sample, however, the estimator will be biased if entrepreneurs on average are more (or less) educated than wage earners.

Our vector of matching variables, S_i, consist of all the variables in Z_{1i} (see Table 4), except the identifying instruments: The dummy-indicators of non-parental entrepreneurs in the family. The matching variables thus include a number of wealth, education and demographic variables, such as, for example *narrow* field of education, age and years of schooling. The matching procedure used is the STATA routine *psmatch2*[17]. We use 1:5 nearest neighbors matching, but the matching algorithm often finds less than 5 matches per entrepreneur. The estimates are not sensitive to the choice of number of neighbors.

Descriptive statistics Table 2 shows that there is a considerable gender imbalance in the population of entrepreneurs. Of the 18,933 incorporated entrepreneurs only 18% (3,474) are women (compared to 43% in the population of full-time employed persons). For the self-employed entrepreneurs, 10,495 out of 44,303 are women (24%). The distribution of education levels is quite similar for men and women, and across types of entrepreneurship. A noticeable exception is the larger fraction of entrepreneurs within higher tertiary education or Phd (18 years or more).

Table 3 shows the survival rates of both self-employed (S) and incorporated (I) entrepreneurs by year of entry into entrepreneurship. Typically, exit from entrepreneurship means that all the entrepreneur's firms are closed down or have become

[17] See http://repec.org/bocode/p/psmatch2.html.

Table 2: **Distribution of educational level among entrepreneurs and in the reference population**

Education level	Incorporated		Self-employed		Population	
	Men	Women	Men	Women	Men	Women
Education level:						
Primary school or lower secondary education (10-)	13 %	15 %	21 %	15 %	20 %	19 %
Post-secondary education (11-13)	50 %	46 %	48 %	39 %	50 %	45 %
Lower tertiary (14-17)	25 %	32 %	18 %	31 %	20 %	30 %
Higher tertiary and Phd (18+)	12 %	7 %	10 %	13 %	8 %	5 %
N (in final sample)	15,459	3,474	33,808	10,495	544,883	417,783

Table 3: **Survival rates of incorporated (I) and self-employed (S) entrepreneurship, by year of entry into entrepreneruship**

Year	Entry-year of entrepreneurship									
	2002		2004		2006		2008		2010	
	I	S	I	S	I	S	I	S	I	S
2002	1.00	1.00								
2003	0.90	0.99								
2004	0.86	0.87	1.00	1.00						
2005	0.78	0.75	0.89	0.96						
2006	0.71	0.69	0.79	0.86	1.00	1.00				
2007	0.67	0.58	0.72	0.70	0.90	0.93				
2008	0.63	0.55	0.65	0.64	0.84	0.85	1.00	1.00		
2009	0.55	0.49	0.58	0.54	0.76	0.72	0.91	0.93		
2010	0.53	0.43	0.55	0.47	0.71	0.60	0.85	0.82	1.00	1.00
2011	0.49	0.38	0.50	0.40	0.65	0.52	0.78	0.70	0.90	0.93

inactive.[18] The table indicates that about 50% of the incorporated entrepreneurship episodes lasts 10 years or more, compared to 40% for the (initially) self-employed entrepreneurs. These differences seem to be quite consistent over time, independent of the start-up year of entrepreneurship. Note that these survival percentages are higher than the usual ones due to the exclusion of necessity entrepreneurs from the sample.

4 Results

The choice to become entrepreneur Table 4 shows the results for the Probit equations of the choice to become an incorporated or self-employed entrepreneur, given the initial state (in 2001) of full time wage employment. Because male and female entrepreneurship and occupational choice decisions have been found to be widely different, we estimate all coefficients separately for males and females. The vector of explanatory variables, Z_{1i}, contains variables regarding the individual's initial condition (in 2001), including region of origin, age, gender, wealth, years of schooling and (narrow) field of education (almost 100 categories). Our main identifying instrument, which in addition to functional form assumptions identifies the return to entrepreneurship, is a dummy variable indicating whether or not at least one non-parent family member is an entrepreneur (self-employed or incorporated) in 2001. We argue and show that this affects the probability of becoming an entrepreneur (positively), but arguably not entrepreneurial earnings directly. Family entrepreneurship has seldom been shown to affect entrepreneurial performance, unlike its great effect on selection into entrepreneurship (Lindquist et al., 2015). Nevertheless, entrepreneurial parents may transmit their entrepreneurial talent, network or experience to their offspring, thereby enhancing their performance as entrepreneurs. Therefore, we cautiously exclude entrepreneurial parents from the set of identifying instruments and consider only non-parental family entrepreneurship. A dummy identifying entrepreneurial parents is included in both the selection and the performance equation.

The results in the table confirm that this instrument is, indeed, a relevant one, with a z-values in the range of 11-17 for the four relevant coefficients (corresponding to combinations of two entrepreneurship types and gender). Other variables with strong associations with entrepreneurship choices are: Years of schooling, Age (inverted U-shaped relation) and Field of education. For incorporated male entrepreneurship, initial wealth is the most important regressor in addition to years of schooling. This is not the case for women. The estimated coefficient of log-wealth in 2001 is twice as large for men than women. For self-employment, initial wealth is hardly of any importance at all. We also note that the positive association between years of schooling and the propensity to become self-employed is much stronger for women than men. The estimated association between having a parent who is self-employed or a corporate owner-manager and (own) entrepreneurship is positive

[18] The exit year is the first year when the entrepreneur is no longer the sole proprietor *or* owner of any active firm which he/she has founded as an entrepreneur. Note that the transition from sole proprietorship to incorporation does not imply exit, but is a continuation of one and the same entrepreneurship episode.

both for men and women, but only significantly so for men. Compared to having a non-parent entrepreneur in the family, these effects are (surprisingly) small.[19]

Given the estimated Probit models of Table 4, it is possible to derive average probabilities to become an entrepreneur during 2002-2011 for all individuals in the reference population in a given category. Some noticeable numbers from such calculations are that: i) A male has on average a three and a half times higher probability of making the transition from employment to incorporated entrepreneurship than a female (2.8 vs 0.8 percent), and more than twice as high a probability of becoming self-employed (6 percent vs 2.5 percent); ii) The only (narrow) field of education where women do not have a substantially lower entrepreneurship probability than men is Medicine (about 25% probability for both sexes); iii) If (at least) one non-parent family member becomes entrepreneur, the average probability that a man (female) in the family also becomes entrepreneur increases from 2.45 percent (0.7 percent) to 4.6 percent (1.0 percent) for incorporated entrepreneurship, and from 7.4 percent (2.5 percent) to 5.9 percent (2.5 percent) for sole proprietorship. Point iii) shows that our identifying instrument is a highly relevant predictor of entrepreneurship, with the possible exception of female self-employment. However, the estimates in Table 4 show that the corresponding coefficient estimate is highly significant also in the latter case.

The returns to entrepreneurship Table 5 depicts the income distribution in the control groups in 2001, i.e., average earnings (in NOK) in each decile, together with ratios of average earnings between the treatment and control groups for the years 2001, 2006 and 2011. While the corresponding control and treatment groups have the same distribution with respect to the matching variables in 2001, they do not have the same earnings distribution. Before they become entrepreneurs, incorporated entrepreneurs have on average higher earnings than individuals in the control group *in all the deciles*; that is, 10-40 percent higher and increasing with higher deciles. On the other hand, self-employed entrepreneurs have almost the same average earnings as the individuals in the control groups, except for the three highest deciles, where they earn 5-10 percent more on average. Thus there appears to be a significant positive selection based on endogenous (unobserved) variables into incorporated entrepreneurship, but not for self-employed entrepreneurs. This selection bias cannot (and should not) be removed by matching, as it is not due to exogenous variables.

The evolution of relative earnings over time is also interesting. Above the fourth decile, the incorporated entrepreneurs (the "treatment group") tend to increase their earnings advantage over the control group over time – but substantially decrease their relative earnings in the two lowest deciles. For the self-employed, there is no evidence of increased relative earnings over time above the fourth decile, but clear evidence of decreasing relative earnings in the three lowest deciles. These figures do not say anything about the earnings profile for any given group of entrepreneurs, as individuals may change earnings decile, or enter and exit entrepreneurship over time. Nevertheless, they strongly indicate that there is a substantial difference in the pecuniary returns to entrepreneurship between the two types of entrepreneurs,

[19] It should here be noted that we are only able to identify living parents in our data set.

Table 4: **Results from estimating the entrepreneurial choice equation (Probit-model)**

Dependent variable: Switch to becoming entrepreneur[1)]	Incorporated		Self-employed	
	Est.	z	Est.	z
Female (dummy)	-0.655	-4.1	-1.710	-17.1
Age				
Male#Age (in 2001)	0.010	3.5	-0.016	-9.1
Female#Age (in 2001)	0.055	10.6	0.039	11.9
$(Age/10)^2$				
Male#$(Age/10)^2$	-0.04	-11.8	-0.003	-1.6
Female#$(Age/10)^2$	-0.09	-13.7	-0.073	-16.9
Years of schooling				
Male#Years of schooling	0.049	27.3	0.015	10.5
Female#Years of schooling	0.033	9.8	0.041	17.5
Dummy for non-parent family- member being entrepreneur:				
Male	0.290	17.1	0.21	14.9
Female	0.242	14.2	0.13	10.8
Dummy for parent being entrepreneur:				
Male	0.031	4.3	0.025	4.2
Female	0.021	1.8	0.003	0.4
log-wealth in 2001				
Male#log-wealth in 2001	0.121	26.2	0.003	2.3
Female#log-wealth in 2001	0.068	12.4	0.012	5.4
No. of individuals in sample	962,666		962,666	

[1)] In addition we have included dummy variables for narrow field of education, country-of-origin, and initial industry of employment (estimates are available upon request)

warranting a formal econometric analysis.

The parameters of main interest in the earnings equation are those of Table 6. The table shows the key results with regard to the estimated treatment effects (ATT) and the control function accounting for selection, ξ_i. Results for the control variables, Z_{2it}, are depicted in Table A1 in the Appendix, with separate estimated effects for men and women.

The parameters of main interest are the estimated average treatment effects on the treated, denoted ATT in Table 6. ATT is defined in (6) for the case with one average treatment effect. However, the results in Table 6 distinguish between average treatment effects along three dimensions: The first dimension is with respect to incorporation or self-employment. The second dimension is with regard to the initial injected equity in the firm, which is either $<$ NOK $200,000$ ("Small incorporated"), or $\geq$ NOK 200,000 ("Large incorporated"), or "All incorporated". The third dimension is with respect to gender: Separate earnings equations and Probit (P) or Ordered probit (O) models of entrepreneurship choice are estimated for men and women. In a third version of our model, no distinction is made between men and women with regard to the treatment effects, but a full set of interaction effects,

Table 5: **Average earnings by decile in the control and treatment groups (relative to the control group)**

Earnings decile[1)]	Control group – earnings (in NOK)		Treatment group – earnings relative to control group					
	Incorp-orated	Self-employed	Incorporated			Self-employed		
	2001	2001	2001	2006	2011	2001	2006	2011
1	160,063	146,808	1.08	0.64	0.36	0.99	0.56	0.55
2	223,427	204,792	1.07	0.94	0.88	1.00	0.79	0.82
3	254,782	239,058	1.08	1.04	1.01	1.01	0.89	0.92
4	279,284	263,299	1.10	1.09	1.09	1.02	0.94	0.97
5	303,300	285,951	1.11	1.13	1.15	1.02	0.97	1.00
6	329,929	309,994	1.14	1.18	1.20	1.03	1.00	1.02
7	363,188	338,869	1.17	1.23	1.25	1.03	1.02	1.03
8	410,298	377,859	1.22	1.30	1.33	1.05	1.04	1.05
9	487,511	442,244	1.28	1.41	1.43	1.07	1.06	1.06
10	685,497	616,882	1.39	1.82	1.72	1.12	1.12	1.09
Share active entrepreneurs[2)]			0 %	41 %	74 %	0 %	58 %	50 %
No. of individuals (N)	71,576	154,956	18,933			44,303		

[1)]The highest and lowest one percent of the earnings observations (both in the treatment and control groups) are excluded in each year [2)]That is, $E(i,t) = 1$

Gender#Y, is included (it is the estimated coefficients of Z_{2it} from this model which are depicted in Table A1 in the Appendix).

The estimated parameter *ATT(All incorporated)* is 0.031 (with a z-value of 8.2), which corresponds to a mere 3.1 percent increase in earnings as a result of becoming entrepreneur when no distinction is made with regard to gender or start-up size. The estimated *ATT(Men#All incorporated)* and *ATT(Women#All incorporated)* are identical. All treatment effects are statistically significant.

If we separate between small and large startups, we see that there is a marked difference in the estimated treatment effects: 0.008 vs 0.102, respectively. The estimated returns to entrepreneurship for individuals who establish small new firms (with less than NOK 200,000 in injected equity), is only slightly positive with a z-value of 1.9. For male entrepreneurs who found small incorporated firms, we get a significantly positive estimate of ATT of 0.012 (with z-value 2.6), whereas the estimate for women is almost identical to zero. For large incorporated start-ups, we find some evidence of smaller ATT for men than for women; with estimates equal to 9.2 and 6.1 percent, respectively. Both are significantly positive. These results indicate that small-scale incorporated entrepreneurs – regardless of gender – to a much lesser degree than large-scale entrepreneurs are motivated by pecuniary returns. Small-scale entrepreneurship is probably a quite close substitute to ordinary employment, and the main motivation for the business venture in this case may not be a high expected pecuniary returns.

Most interestingly, we find a significantly negative return to self-employment. The estimated ATT(*Self-employed*) is -0.065, i.e. a negative return of more than 6 percent (which is highly significant, with a z-value of -34.7). The gender-specific

ATT-estimates are respectively -0.067 for men and -0.099 for women – both are highly significant. Thus self-employment seems to be even less motivated by the prospect of pecuniary returns than small scale incorporation.

The coefficient of the control function $\xi_i(e)$ in Table 6 is a measure of the degree of selection by *absolute advantage* into entrepreneurship (cf., the discussion in Section 2). The positive estimates of around 0.06 (z-values above 40) for men and 0.04 (z-values above 14) for women, tell us that selection by absolute advantage strongly characterizes the selection into incorporated entrepreneurship. Persons who become entrepreneurs have cet. par. much higher earnings potential than wage-earners *irrespective* of their choice to become entrepreneurs. As a consequence, the average observed earnings differentials between entrepreneurs and non-entrepreneurs (OD) – everything else equal – is higher than the ATT (cf. (13). If we exclude the control function $\xi_i(e)$ from the model, the estimated *ATT(All startups#incorporated)* increases with about 2 percentage points for all the different ATT-effects of incorporated entrepreneurship. In contrast we find no evidence of selection by absolute advantage into self-employment. In fact, the estimated coefficient of the control function, $\xi_i(e)$, is negative for male self-employed entrepreneurs (-0.006), and even slightly significant, with a z-value of -5.8. We conclude that there is an ambiguous effect of general income ability on the propensity to become a self-employed entrepreneur, but that the magnitude of the effect is close to zero.

Table 7 displays the results for the second order moments. In general, the standard deviation, denoted $\text{sd}(\cdot)$, of log-earnings for active entrepreneurs (that is, when $E(i,t) = 1$), are much higher than in the control groups ($E_i = 0$); 0.55 and 0.68 for incorporated and self-employed entrepreneurs, respectively, compared to 0.41 and 0.43, respectively, in the corresponding control groups. As expected, the earnings dispersion for large incorporated entrepreneurs are larger than for small ones (standard deviations equal to 0.70 vs 0.64). These differences can almost entirely be attributed to being an active entrepreneur, as $\text{sd}(u_{it})$ and $\text{sd}(\varepsilon_{2i})$ are very similar across the treatment ($E_i > 0$) and control groups ($E_i = 0$). Moreover, the estimates of the autoregressive coefficient ϕ in Table 7 (equal to 0.59 and 0.55) reveal a high degree of autocorrelation in the error term u_{it}.

Table 6: **Estimated average treatment effects on the treated (ATT) according to gender, control function (yes/ no), startup size (large/small) and type of entrepreneur (incorporated or self-employed)**

Dependent variable: log-earnings								
Startup size and type:	Small[1)] incorp.		Large incorp.		All incorp.		Self-employed	
Control function included:	Yes	No	Yes	No	Yes	No	Yes	No
Control function derived from:[2)]	O	–	O	–	P	–	P	–
All individuals								
ATT	0.008	0.027	0.102	0.125	0.031	0.051	-0.065	-0.067
(z-value)	(1.9)	(6.6)	(18.4)	(22.7)	(8.2)	(13.7)	(-34.7)	(-36.1)
$\xi_i(e)$	0.057	–	0.057	–	0.055	–	-0.004	–
(z-value)	(44.0)	(–)	(44.0)	(–)	(41.9)	(–)	(-4.5)	(–)
No. of entrepreneurs	12,597		6,336		18,933		44,303	
Men								
ATT	0.012	0.031	0.092	0.115	0.031	0.051	-0.060	-0.062
(z-value)	(2.6)	(6.8)	(15.5)	(19.4)	(7.7)	(12.7)	(-28.7)	(-30.2)
$\xi_i(e)$	0.059	–	0.059	–	0.058	–	-0.006	–
(z-value)	(40.1)	(–)	(40.1)	(–)	(38.7)	(–)	(-5.8)	
No. of entrepreneurs	10,084		5,375		12,597		33,808	
Women								
ATT	-0.004	0.009	0.061	0.078	0.032	0.039	-0.099	-0.098
(z-value)	(-0.4)	(0.9)	(5.6)	(7.2)	(3.4)	(4.7)	(-22.3)	(-22.6)
$\xi_i(e)$	0.040	–	0.040	–	0.046	–	0.000	
(z-value)	(14.3)	(–)	(14.3)	(–)	(16.7)	(–)	(0.1)	
No. of entrepreneurs	5,375		961		6,336		10,495	

[1)] $< 200,000$ NOK in initial injected equity [2)]Ordered probit (O) or Probit (P)

Table 7: **Estimates of parameters pertaining to the second order moments of the earnings equation**

Dependent variable: log-earnings	Small[1)]	Large	All	Self-
Startup size and type:	incorp.	incorp.	incorporated	employed
Control function derived from:[2)]	O	O	P	P
sd(log-earnings)|active entrepr, Z_2)[3)]	0.64	0.70	0.68	0.55
sd(log-earnings)|control group, Z_2)[3)]	0.41	0.41	0.41	0.43
Residual (u_{it}):				
AR-coefficient (ϕ)	0.59	0.59	0.59	0.55
sd(u_{it})	0.32	0.32	0.32	0.35
Random effect (ε_{2i})				
sd(ε_{2i}|treatment group)	0.32	0.34	0.32	0.29
sd(ε_{2i}|control group)	0.26	0.26	0.26	0.26

[1)] $< 200,000$ NOK in initial injected equity [2)]Ordered probit (O) or Probit (P)
[3)]Standard deviation of log-earnings for active entrepreneurs ($Ent(i,t) = 1$), and in the control group ($E_i = 0$) conditional on control variables (Z_2)

The estimated effects of the control variables, Z_{2it}, are displayed in the Appendix (see Table A1). These are in general of a sign and magnitude as one would expect from the empirical wage equation literature. Years of schooling and years of experience are the most significant explanatory variables in the model. An additional year of schooling increases earnings by an estimated 6-7 percent (slightly more for women than men). Years of experience has an inverted U-shaped effect on earnings. We see that there is a significant negative impact from the interaction variable *financial crisis*×*entrepreneur* (*financial crisis* is a dummy for 2008-2009) and that initial wealth (in 2001) has a positive impact on earnings in all years. For a given level of schooling, the highest earnings are observed in the education fields Social science and law and Business and administration. There are few notable differences between males and females or across entrepreneurship types, with respect to the impact of exogenous variables. In particular, the trend in log-earnings from 2001–2011 (not shown in the table) is almost identical for men and women, with about 4 percent nominal annual increase on average. The effect of having a parent who is self-employed or a corporate owner-manager is ambiguous. Regardless of entrepreneurship type, the effect on earnings is positve for men (0.02), but negative for women (-0.05).

5 Conclusion

In this paper we have reconsidered the so-called 'entrepreneurial earnings puzzle', i.e., the finding – by most studies – of zero or negative returns to entrepreneurship. Our analyses have been based on two pillars: data improvements and model improvements. First, we have had the advantage of registry data comprising the whole Norwegian labor population and all firms established in the period 2002-2011. Second, these data have allowed us to identify both sole proprietors and owners of incorporated firms and their ownership shares. Third, we have observed very detailed measures of employment and entrepreneurship incomes in the registry. Our

rich data have also allowed us to propose a uniform measure of (pretax) earnings for all individuals in the sample, whether self-employed, wage-employed, unemployed or owner-managers of incorporated firms. It consists of the sum of labor income from wages and self-employment, work-related cash transfers and owner income from incorporated firms in which the individual is an entrepreneur.

Our analyses have focused on 'offensive entrepreneurship', i.e. the transition from full-time wage employment to entrepreneurship. We have also distinguished between unincorporated (self-employed) and incorporated entrepreneurs, and – within the latter group – between 'large' and 'small' start-ups. When estimating the returns to entrepreneurship ('the average treatment effects on the treated'), our model has enabled us to take into account that the choice to become an entrepreneur is endogenous with respect to earnings prospects.

The main bulk of results found in entrepreneurship journals pertain to self-employment and usually find zero or negative returns. In line with these findings, we found that the average return to entrepreneurship is significantly negative for individuals entering entrepreneurship through self-employment. On the other hand we found that persons who become entrepreneurs by establishing relatively large incorporated firms, increase their earnings by 10 percent on average by becoming entrepreneurs. The latter results are more in line with studies including 'high end' entrepreneurs such as (variants of) incorporated entrepreneurship (Berglann et al., 2011; Hvide, 2009; Hvide and Panos, 2014; Levine and Rubinstein, 2013), or even billionaire entrepreneurs (Henrekson and Sanandaji, 2014). All these studies find positive returns to entrepreneurship.

Nevertheless, some of our results are surprising in view of comparable analyses on registry data, especially Berglann et al. (2011), who find that entrepreneurs are *overall* generously rewarded in Norway. One explanation of this discrepancy may be that they identify different entrepreneurs than us by effectively conditioning on future outcomes (i.e. that the entrepreneurship earnings is the 'most important' source of income). However, by conditioning on future outcomes, one does not capture the full *ex ante* risk and reward of the transition from wage employment to entrepreneurship.

One must be careful about drawing too stark policy implications of our results. While OECD (2003; 2005) considers entrepreneurship as an important source of economic growth and innovation in the economy, our results indicate that there may not be much to gain in economic terms for the individual entrepreneur. In a country like Norway – with a high employment rate among both men and women – there may be less to gain both for the individual entrepreneur and society as a whole through small-scale entrepreneurship than in countries where there are more unemployed resources that may be mobilized into the labour market through self-employment. Our analyses indicate that, at least to the individual entrepreneur, the ordinary labor market may pay off just as well – at much less risk – than managing one's own business.

References

[1] Aastebro, T. (2012): The returns to entrepreneurship. In D. Cumming (ed.): *The Oxford Handbook of Entrepreneurial Finance*, Oxford: Oxford University Press.

[2] Aastebro, T. and J. Chen (2014): The entrepreneurial earnings puzzle: mismeasurement or real? *Journal of Business Venturing*, 29, 88-105.

[3] Benz, M. and B. Frey (2008): Being independent is a great thing: Subjective evaluations of self-employment and hierarchy, *Economica*, 75, 362-383.

[4] Berglann, H., E.R. Moen, K. Roed, and J.F. Skogstrøm (2011): Entrepreneurship: Origins and Returns. *Labour Economics*, 18, 180-193.

[5] Berkhout, P., J. Hartog, and van C.M. Praag (2011): It's the Opportunity Cost, stupid! How Self-employment responds to financial incentives of return, risk and skew, Tinbergen Institute Discussion Papers, 11-165/3, Tinbergen Institute.

[6] Blanchflower, D.G. and A.J. Oswald (1998): What makes an entrepreneur? *Journal of Labor Economics*, 16, 26-60.

[7] Carter, S. (2011): The rewards of entrepreneurship: exploring the incomes, wealth, and economic well-being of entrepreneurial households. *Entrepreneurship Theory and Practice*, 35, 39–55.

[8] Dagsvik, J.K., T. Hægeland, A. Raknerud (2011): Estimating the Returns to Schooling: A Likelihood Approach Based on Normal Mixtures. *Journal of Applied Econometrics*, 26, 613–640.

[9] Dushnitsky, G. (2010): Entrepreneurial optimism in the market for technological invention, *Organization Science*, 21, 150-167.

[10] Feldman, N.E. and J.B. Slemrod (2007): Estimating tax noncompliance with evidence from unaudited tax returns, The Economic Journal, 117(518), 327-352.

[11] Fjærli, E., D. Iancu and A. Raknerud (2013): Facts about entrepreneurship in Norway. Who become entrepreneurs and how do they perform? Reports, 52/2013, Statistisk sentralbyrå.

[12] Garen J. 1984. The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable. *Econometrica* 52: 1199–1218.

[13] Gourieroux, C. and A. Montfort (1995). *Statistics and Econometric Models*. Cambridge University Press, Cambridge.

[14] Hayward, M., D. Shepherd and D. Griffin (2006): A hubris theory of entrepreneurship. *Management Science,* 52, 160-172.

[15] Heckman, JJ. 1979. Sample Selection Bias as a Specification Error. *Econometrica* 47: 153–162.

[16] Heckman, J. and Guilherme L. Sedlacek (1990): Self-Selection and the Distribution of Hourly Wages. *Journal of Labor Economics*, Vol. 8, No. 1, Part 2: Essays in Honor of Albert Rees.

[17] Henrekson, M. and T. Sanandaji (2014): Small business activity does not measure entrepreneurship. Proceedings of the National Academy of Sciences, 111(5), 1760-1765.

[18] Hamilton, B. H. (2000): Does entrepreneurship pay? An empirical analysis of the returns to self-emplyment. *Journal of Political Economy*, 108, 604-631.

[19] Holm, H., S. Opper and V. Nee (2013): Entrepreneurs under uncertainty: an economic experiment in China. *Management Science*, 59, 1671-1687.

[20] Hurst, E. and A. Lusardi (2004): Liquidity constraints, household wealth and entrepreneurship. *Journal of Political Economy*, 112, 319-347.

[21] Hurst, E., G. Li and B. Pugsley (2013): Are household surveys like tax forms: Evidence from income underreporting of the self employed. *Review of Economics and Statistics*, 96, 19-33.

[22] Hvide, H.K. (2009): The Quality of Entrepreneurs. *Economic Journal*, 119, 1010-1035.

[23] Hvide, H. and J. Møen (2010): Lean and hungry or fat and content? Entrepreneurs' wealth and start-up performance. *Management Science*, 56, 1242-1258.

[24] Hvide, H. and G. Panos (2014): Risk tolerance and entrepreneurship. *Journal of Financial Economics*, 111, 200-223.

[25] Koudstaal, M., R. Sloof and C.M. van Praag (2014): Risk, uncertainty and entrepreneurship: evidence from a large lab-in-the-field experiment. Unpubished manuscript.

[26] Levine, R. and Y. Rubinstein (2013): Smart and illicit: Who becomes an entrepreneur and does it pay? NBER Working Pap No. 19276. National Bureau of Economic Research, Cambridge, MA.

[27] Lindquist, M., J. Sol and C.M. Van Praag (2015): Why do entrepreneurial parents have entrepreneurial children? *Journal of Labor Economics*, Forthcoming.

[28] Lowe, R. and A. Ziedonis (2006): Overoptimism and the performance of entrepreneurial firms. *Management Science*, 52, 173-186.

[29] Luger, M. I. and J. Koo (2005): Defining and tracking business start-ups. *Small Business Economics,* 24, 17-28.

[30] McCraffrey, M. (2014): On the theory of entrepreneurial incentives and alertness. *Entrepreneurship Theory and Practice*, 38, 891–911.

[31] Myers, S.C . and N.S. Majluf (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, 13, 187–221.

[32] Manski, C. F. and L. R. Lerman (1977): The estimation of choice probabilities from choice based samples. *Econometrica*, 45, 1977-1988.

[33] Nicolaou, N., S. Shane, L. Cherkas, J. Hunkin and T.D. Spector (2008): Is the tendency to engage in entrepreneurship genetic? *Management Science*, 54, 167-179.

[34] OECD (2003) The Sources of Economic Growth in OECD Countries. Paris: OECD

[35] OECD (2005): Understanding Economic Growth. Paris: OECD.

[36] Parker, S. C. (2004). *The economics of self-employment and entrepreneurship.* Cambridge University Press.

[37] Parker, S.C. (2009): *The Economics of Entrepreneurship.* Cambridge: Cambridge University Press.

[38] van Praag, C. M. and S. C. Parker (2006): Schooling, capital constraints, and entrepreneurial performance. The endogenous triangle, *Journal of Business & Economic Statistics*, 24, 416-431.

[39] Rosenbaum, P.R. and D.B. Rubin (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 4.1

[40] van der Sluis, J., M. van Praag and A. van Witteloostuijn (2004): Comparing the returns to education for entrepreneurs and employees. Tinbergen Institute Discussion Papers, 104/3.

Appendix: Supplementary materials

Proof of Theorem 1.

$$\begin{aligned} & E(\varepsilon_2|E=e) \\ = & \frac{1}{P(E=e))} E\left(\varepsilon_2 1\{\mu_e - Z_1\gamma_1 < \varepsilon_1 \leq \mu_{e+1} - Z_1\gamma_1\}\right) \\ = & \frac{1}{P(E=e)} \int_{\mu_e - Z_{1i}\gamma_1}^{\mu_{e+1} - Z_{1i}\gamma_1} \varepsilon_2 \phi(\varepsilon_2) d\varepsilon_2 \\ = & \frac{\phi\left(\mu_e - Z_1\gamma_1\right) - \phi\left(\mu_{e+1} - Z_1\gamma_1\right)}{\Phi\left(\mu_{e+1} - Z_{1i}\gamma_1\right) - \Phi\left(\mu_e - Z_1\gamma_1\right)} \\ = & \xi(e) \end{aligned}$$

This completes the proof. ■

Table A1: **Parameter estimates of control variables in earnings equation**

Dependent variable: log-earnings[1)]	Incorporated				Self-employed			
	Est.	z	[95%	CI]	Est.	z	[95%	CI]
Interaction effects males:								
Years of schooling	0.06	123.5	0.06	0.06	0.05	138.2	0.05	0.05
Years of experience	0.04	178.1	0.04	0.04	0.04	238.6	0.04	0.04
(Years of experience/10)^2	-0.08	-176.8	-0.08	-0.08	-0.07	-235.1	-0.07	-0.07
Field of education:								
General programmes	0	(ref)						
Humanities	-0.08	-12.4	-0.09	-0.07	-0.07	-18.6	-0.08	-0.07
Teacher training	0.00	-0.6	-0.02	0.01	0.00	0.5	-0.01	0.01
Social science and law	0.07	10.4	0.05	0.08	0.06	12.7	0.05	0.07
Business and adm.	0.13	41.3	0.12	0.13	0.11	44.5	0.10	0.11
Natural scienes	0.07	29.6	0.06	0.07	0.07	46.1	0.07	0.07
Health	0.04	6.9	0.03	0.05	0.07	18.4	0.06	0.08
Transport services.	0.10	22.6	0.09	0.11	0.07	23.7	0.07	0.08
Entrepreneur(dummy)#crisis	-0.05	-19.7	-0.06	-0.05	-0.02	-12.2	-0.02	-0.02
log-wealth in 2001	0.04	55.3	0.04	0.04	0.03	79.8	0.03	0.03
Parent-entrepreneur (dummy)	0.02	5.7	0.01	0.03	0.02	7.61	0.02	0.03
Interaction effects females:								
Dummy for being female	0.15	3.9	0.07	0.22	0.22	12.2	0.18	0.25
Years of schooling	0.07	58.9	0.07	0.07	0.05	86.9	0.05	0.05
Years of experience	0.02	50.6	0.02	0.03	0.03	89.7	0.03	0.03
(Years of experience/10)^2	-0.04	-42.1	-0.04	-0.03	-0.04	-75.3	-0.04	-0.04
Field of education:								
General programmes	-0.06	-2.6	-0.11	-0.02	0.01	1.0	-0.01	0.04
Humanities	-0.09	-3.9	-0.14	-0.05	0.00	0.1	-0.02	0.02
Teacher training	-0.05	-2.2	-0.10	-0.01	0.06	4.9	0.04	0.09
Social science and law	0.03	1.1	-0.02	0.08	0.12	9.3	0.09	0.14
Business and adm.	0.04	1.7	-0.01	0.09	0.10	8.6	0.08	0.13
Natural scienes	0.03	1.1	-0.02	0.07	0.12	10.0	0.10	0.15
Health	-0.02	-0.8	-0.07	0.03	0.07	5.9	0.05	0.10
Transport services.	-0.02	-0.6	-0.06	0.03	0.03	2.4	0.01	0.06
Entrepreneur(dummy)#crisis	-0.01	-3.5	-0.02	-0.01	-0.04	-15.5	-0.04	-0.03
log-wealth in 2001	0.04	24.4	0.03	0.04	0.03	36.9	0.02	0.03
Parent-entrepreneur (dummy)	-0.05	-9.5	-0.06	-0.04	-0.05	-12.6	-0.6	-0.04

[1)] Dummies for year and country-of-origin are included in the estimated model, but not shown

Statistics Norway

Postal address:
PO Box 8131 Dept
NO-0033 Oslo

Office address:
Akersveien 26, Oslo
Oterveien 23, Kongsvinger

E-mail: ssb@ssb.no
Internet: www.ssb.no
Telephone: + 47 62 88 50 00

ISSN: 1892-753X

Statistisk sentralbyrå
Statistics Norway

Design: Siri Boquist