DISCUSSION PAPER SERIES

IZA DP No. 7206

Risk Tolerance and Entrepreneurship

Hans K. Hvide
Georgios A. Panos

February 2013

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Risk Tolerance and Entrepreneurship

Hans K. Hvide
University of Bergen, CEPR, IZA
and University of Aberdeen

Georgios A. Panos
University of Stirling,
CER and CELMR

Discussion Paper No. 7206
February 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

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

IZA Discussion Paper No. 7206
February 2013

ABSTRACT

Risk Tolerance and Entrepreneurship*

A tradition from Knight (1921) argues that more risk tolerant individuals are more likely to become entrepreneurs, but perform worse. We test these predictions with two risk tolerance proxies: stock market participation and personal leverage. Using investment data for 400,000 individuals, we find that common stock investors are around 50 percent more likely to subsequently start up a firm. Firms started up by stock market investors have about 25 percent lower sales and 15 percent lower return on assets. The results are similar using personal leverage as risk tolerance proxy. We consider alternative explanations including unobserved wealth and behavioral effects.

JEL Classification: L26

Keywords: entrepreneurial entry, entrepreneurial performance, firm entry, firm performance, firm productivity, firm survival, overconfidence, risk aversion, risk tolerance, stock market participation

Corresponding author:

Hans K. Hvide
University of Bergen
Department of Economics
Fosswinckels gate 14
5007 Bergen
Norway
E-mail: hans.hvide@econ.uib.no

* We thank an anonymous referee for valuable comments, and Rob Fairlie, Terry Odean, Shackar Kariv, Francisco Peres-Gonzales, Mirjam van Praag, Uwe Sunde, Bertil Tungodden, Arild Aakvik, Karl Ove Aarbu, and seminar participants for discussion and suggestions. This work has received financial support from the Norwegian Research Council under grant number 178944. We thank Verdipapirsentralen (The Norwegian Central Securities Depository) and Visma for providing data. Many thanks to Øystein Hungerholt and Tarik Turkanovic for assistance with the data.

1 Introduction

While corporate finance theory takes the existence of firms as given, financial economists are increasingly interested in where firms come from, and which factors affect the inception and growth of new firms (e.g., Rajan, 2012). In this paper we provide new evidence on the hypothesis from Knight (1921) that less risk averse individuals are more likely to start up a firm. We also examine the relation between entrepreneur risk preferences and firm performance. The data comes from Norway and comprises longitudinal information on all firms started up in that country over an extended period.

Much work on entrepreneurial firms has focused on access to credit, and little is known about the role of risk preferences.[1] This question is likely to be important because entrepreneurial households appear not to be able to diversify the risk of their business well (e.g., Moskowitz & Vissing-Jorgensen, 2002, Bitler et al., 2005). In the theoretical models of Kanbur (1979) and Kihlstrom & Laffont (1979), less risk averse individuals become entrepreneurs, and more risk averse individuals become workers. An additional hypothesis that follows from theory is that less risk averse individuals are likely to perform worse as entrepreneurs. In this paper we attempt to test these hypotheses.

Do risk preferences play a role in the origin and growth of firms? Our empirical approach is to construct risk tolerance proxies based on individual investment histories, and test whether there is a link between "revealed risk tolerance" and the decision to start up a firm.[2] Using a similar approach, we test whether there is a link between the revealed risk tolerance of entrepreneurs and their firm's growth, profitability and survival. We employ four proxies for risk tolerance. We use stock market participation and personal leverage. Relative to investments in government bonds or savings accounts, common stock

[1]A tradition in management uses cross-sectional survey evidence to study risk attitudes of entrepreneurs versus non-entrepreneurs. Generally they find no differences (Shane et al., 2003, Wu & Knott, 2006), but see also Cramer et al. (2002) and Dohmen et al. (2009). Caggese (2012) shows that investments of firms with a dominant owner that is also the manager ("entrepreneurial firms") are more sensitive to increased uncertainty than other firms, consistent with entrepreneurial risk aversion.

[2]The literature on eliciting preferences in economics usually involves experiments with stakes that have little impact on lifetime resources (e.g., Choi et al., 2007, Dohmen et al., 2009), or answers to hypothetical survey questions (e.g., Barsky et al., 1997). In contrast, we use proxies that are based on large-stake past investment decisions in a natural setting. See Malmendier & Tate (2008) for a similar approach to elicit and test for the importance of managerial overconfidence.

investments are indisputably risky; for example, the average standard deviation of yearly returns for individual stocks in the U.S. is about 25 percent (Campbell et al., 2001).[3] A higher personal leverage, defined as a fraction of income or wealth to debt, makes an individual more likely to experience financial distress and bankruptcy. This measure has no direct link to the stock market, and complements those results. As additional risk tolerance proxies, we use the fraction of wealth invested in the stock market (e.g., Merton, 1969) and the volatility of stock portfolio returns at the individual level.

We use data from Norway which covers all new firms incorporated between 2000 and 2007. Covering the population of new firms means that the vast majority of firms in our database are small. The advantage of this approach is that it will not be subject to selection biases commonly encountered in the literature that uses "tip-of-the-iceberg" datasets. For each firm, the data identifies the initial owners. The dataset also contains details on the investment history of *all* Norwegian individuals prior to 2000, including yearly asset balance and investments in common stocks. In our sample, consisting of males that are fully employed in 1993-1994, there are some 400,000 individuals, of which about 70,000 invest in common stocks between 1995 and 1999 and about 6,300 become entrepreneurs, defined as having a majority stake in a new firm outside portfolio investment.

We find that all four proxies are strongly related to starting up a firm. Individuals are more likely to become entrepreneurs if they participate in the stock market or have a high personal leverage. In addition, conditional on participating in the stock market, individuals are more likely to become entrepreneurs if they invest a higher fraction of their wealth in the stock market or have a more volatile stock portfolio return. The economic magnitudes are large. For example individuals that invest in common stocks are 50 percent more likely to become entrepreneurs. For males around thirty years old the estimated effect is over 70 percent. The results are similar using personal leverage (or the other risk tolerance proxies).

The main identification assumption is that unobserved factors that affect the risk tolerance proxies are uncorrelated with the entry decision. In all our regressions, we

[3]The corresponding Norwegian figure is about 35 percent. Many individuals invest in more than one stock, reducing portfolio risk, but in too few stocks to be diversified properly. The average standard deviation of yearly portfolio returns in Norway is about 30 percent in the sample period.

control for income, age, wealth, and education, in addition to work background. Wealth is a particularly important control, as it can be expected to affect both investments and entrepreneurial activity. We use several different measures of wealth and obtain very similar results. For example, we use gross wealth, net wealth, and net capital income. Moreover, any remaining unobserved differences in wealth are unlikely to explain all four of our risk tolerance measures. Our results also hold when excluding investors whose portfolio returns exceed the risk-free rate.

Stock ownership could increase an individual's productivity as entrepreneur due to learning about capital markets or on how to collect business-relevant information, and thereby lead to higher entry rates. If selection on risk tolerance drives our results, in contrast, we expect that stock market participants perform worse as entrepreneurs. The reason, derived in our theoretical analysis, is that less risk averse individuals would be willing to accept lower expected entrepreneurial returns for given risk (see also Kanbur, 1979, and Kihlstrom & Laffont, 1979).[4]

We analyze the relation between founder risk tolerance and startup performance using longitudinal accounting data that covers each startup from the first year onwards. We find that firms started up by more risk tolerant individuals are less profitable, grow less, and have lower survival rates. For example, firms started up by stock market investors have about 13 percent lower return on assets. The results hold up independently of whether we use stock market participation or personal leverage as a proxy for risk tolerance. All our performance estimations control for year, industry and size effects, in addition to the sociodemographic controls used in the entry analysis.

The empirical literature suggests that behavioral traits influence asset allocation. For example, individual investors typically own too few stocks to be well diversified, and they trade excessively (e.g., Odean, 1999, Barber & Odean, 2000). Odean (1998) and Barber & Odean (2000) suggest that excess trading is driven by overconfidence, a trait that is likely correlated with entrepreneurial entry (e.g., Landier & Thesmar, 2009). We ask

[4] As discussed further in the text, if projects with higher mean returns also tend to have a higher variance, then more risk-tolerant entrepreneurs would select projects whose variance is higher, and whose mean returns could be higher. In the data, the variance of operating returns on assets (OROA) is only weakly higher for individuals that are identified as more risk tolerant, so that a constant variance assumption appears to be a reasonable approximation.

whether there is a relation between risk tolerance and entrepreneurship after taking overconfidence into account. The empirical analysis suggests that overconfidence, as measured by trading intensity, is associated positively with the propensity to become an entrepreneur (supporting the notion that overconfidence can explain entrepreneurial entry), but appears unrelated to the role played by risk tolerance. We also attempt to account for the possibility that a preference for sensation-seeking (Grinblatt & Keloharju, 2009) drives our results, by including a control for the engine size of one's car (relative to car size). Including a control for engine size has no impact on the estimated relationship between risk tolerance and entrepreneurship.

The paper connects to the literature in several ways. First we contribute to the debate over the origin of firms, and which factors can explain their inception and growth. Much literature has focused on financial constraints (e.g., Evans & Jovanovic, 1989, Hvide & Moen, 2010, Robb & Robinson, 2012, Andersen & Nielsen, 2012) or the role of business angels and venture capitalists (e.g., Kerr et al, 2011). We focus on the individuals that start up these firms, and find support for the Knight (1921) conjecture that less risk averse individuals self-select into entrepreneurship.[5] Using a sample of venture capital financed firms that undergo IPOs, Kaplan et al. (2012) find that the management team frequently changes while the line of business is almost constant. Our results complement Kaplan et al. (2012) by pointing out that individuals, or more precisely individual traits, do seem important to understand the inception and growth of very young firms. Our research also complements the Lerner & Malmendier (2011) finding that learning about other individuals' entrepreneurial experiences decreases entry rates but improves performance. We find that risk aversion has a similar effect: more risk averse individuals have lower entry rates but superior performance. Our findings also contribute to the literature on firm productivity. Economists have shown that large and persistent differences in productivity across firms exist even after taking into account geographical, industry and firm age differences (e.g., Syverson, 2011). We point out that a factor missing in this literature – individual entrepreneurs – can explain some of the heterogeneity for young firms. For

[5] A literature considers the effects of individuals' human capital and work background (e.g., Lazear, 2005, Hvide, 2009) and their networks (e.g., Gompers et al, 2005, Nanda & Sorensen, 2010). A large literature, reviewed by Parker (2009), focuses on the self-employed.

example, our findings suggest that a possible reason why many young businesses fail in the beginning (e.g., Audretsch & Mahmood, 1995) could be because risk tolerant individuals with relatively poor ideas choose entrepreneurship.[6] Finally, our work also relates to a recent literature that focuses on the role of individuals and their traits, as opposed to institutional factors, in shaping economic outcomes (e.g., Bertrand & Schoar, 2003, Bennedsen et al., 2007, Jones & Olken, 2005, Kaplan et al, 2012, Malmendier & Tate, 2008). While this literature focuses on the role of national leaders and CEOs in shaping policies, we focus on the link between individuals and the origin and growth of firms.

The remainder of this paper is organized as follows. Section 2 provides a theoretical framework. Section 3 analyzes entrepreneurial entry, and Section 4 analyzes entrepreneurial performance. Section 5 analyzes behavioral issues and Section 6 concludes.

2 Theoretical backdrop

Here we present a simple model that forms the basis of the empirical analysis.

2.1 Entry

Theories of firm entry that encompass the incentives of entrepreneurs suggest that personal wealth, opportunity cost of human capital, and risk tolerance are the main determinants (Evans & Jovanovic, 1989; Kanbur, 1979; Kihlstrom & Laffont, 1979, Lucas, 1978). Let X_i be a stochastic vector of observable characteristics for individual i, and let the scalar r_i be a measure of individual i's risk aversion. By risk tolerance we mean $1/r_i$. We assume that r_i is stochastic with mean μ and standard deviation σ_r. Skipping i subscripts, we assume an individual starts up a firm if $e = 1$, where

$$e = \mathbf{1}(\alpha_E + X\beta_E - r\delta_E - \epsilon_E). \tag{1}$$

[6] Related to the theoretical literature on industry evolution, our findings inform standard models, where entering firms are assumed to be ex-ante homogenous (e.g., Jovanovic, 1982, Hopenhayn, 1992, Asplund & Nolcke, 2006).

1(.) is an indicator function that equals one if the expression inside the brackets is positive, zero if not. The term ϵ_E is iid random with $E(\epsilon_E|X, r) = 0$ and $Var(\epsilon_E|X, r) = \sigma_E^2 < \infty$. The elements of X are allowed to be correlated, and we also allow for correlation between X and r.[7] The expression inside the brackets of (1) can be interpreted as the expected utility differential between starting up a firm and alternative occupations, where $r\delta_E$ is the "risk-cost" of entrepreneurship. The risk-cost is large if either risk aversion is high (a high r), or if the returns from entrepreneurship are much more risky than alternative occupations, i.e., δ_E is high. With the exception of the $r\delta_E$ term, empirical models such as (1) are well established in the empirical literature on entrepreneurial entry (Hurst & Lusardi, 2004). Appendix B shows that (1) can be derived from an underlying latent utility model with constant absolute risk aversion.

2.2 Proxies for risk tolerance

Since the risk-preference parameter r is not observable, one cannot estimate (1) directly. Therefore, we use stock market participation as a proxy for r, among other proxies. Asset allocation theory suggests risk tolerance is the only individual characteristic that determines stock market participation if markets are complete (Merton, 1969; Mossin, 1968; Samuelson, 1969); a less wealthy individual's portfolio should be a scaled-down version of a wealthy individual's portfolio. Under the realistic assumption that markets are incomplete due to undiversifiable labor income risk, theory identifies a role for determinants of future labor income when determining portfolio choice (Heaton & Lucas, 1997; Cocco et al., 2005). Entering the stock market involves fixed costs (Vissing-Jorgensen, 2002; Campbell, 2006), including acquiring information, becoming aware of stock market opportunities (Guiso & Japelli, 2005), and possibly psychic costs due to limited trust in the stock market (Guiso et al., 2008).[8] With fixed costs of entry, wealth plays an important role in determining stock market participation.

[7] Such correlation arises for several reasons. Highly educated individuals may be more risk tolerant (Dohmen et al., 2009); precautionary savings could lead risk-averse households to accumulate more wealth (Buera, 2008), or risk aversion may depend on wealth.

[8] The Norwegian population is homogenous and the financial markets well-developed; we expect trust to be uniformly high in the context of our data.

We assume that an individual prefers to invest in common stocks if $I = 1$, where

$$I = \mathbf{1}(\alpha_I + X\beta_I - r\delta_I - \epsilon_I). \tag{2}$$

ϵ_I is iid random with $E(\epsilon_I|X, r) = 0$ and $Var(\epsilon_I|X, r) = \sigma_I^2 < \infty$, and $r\delta_I$ reflects the risk-cost of stock market participation.[9] With the exception of the term $r\delta_I$, (2) is a standard regression model in the stock market participation empirical literature (Hong et al., 2004; Campbell, 2006). An implicit assumption from (1) and (2) is that the risk preference parameter r is stable over time and across decision problems. This assumption is debatable but consistent with panel evidence from Andersen et al., (2008) and Sahm (2007), and with cross-sectional evidence from Barseghyan et al. (2011) and Dohmen et al. (2009).

Our other proxies for risk tolerance can be analyzed in a similar way. We assume that the fraction of wealth put in the stock market, and personal leverage, are determined by wealth and other sociodemographic characteristics in addition to risk tolerance,

$$I = \alpha_I + X\beta_I - r\delta_I - \epsilon_I. \tag{3}$$

Note that the risk tolerance proxy I is here a continuous variable.

2.3 Estimation of entry

We use I as a proxy for r and estimate the modified entry equation

$$e = \mathbf{1}(\alpha + X\beta + I\gamma - \epsilon). \tag{4}$$

Using (4), we estimate $dE(e|X, I)/dI$, denoted by e_I, i.e., how much the entrepreneurship rate changes with a change in risk tolerance proxy I.[10] Estimating (4) by ordinary least squares yields the same estimate of e_I as a two-step procedure where first I is regressed on

[9] It is convenient but unessential that the sociodemographic vector in (2) is the same as in (1). Kumar (2009) argues that a fraction of individual investors has a preference for gambling, i.e., are risk lovers. Risk-loving preferences are captured by r being negative in (2).

[10] This is a slight abuse of notation as the derivative is not defined under (2). For that case, the notation $dE(e|X, I)/dI$ should be taken as a shorthand for $E(e|X, I = 1) - E(e|X, I = 0)$.

X using ordinary least squares, and then residuals from this regression - the unexplained part of I - are proxies for risk tolerance. The vector X in (4) therefore serves the dual role of controlling for other factors that explain I and for other factors that explain entrepreneurial entry.

Identification of the risk tolerance effect relies on the following assumption: conditional on predetermined demographic characteristics, both variation in risk tolerance and revealed risk-tolerance are independent of unobserved determinants of entry. Formally, the identification assumption is that $Cov(\epsilon_E, \epsilon_I) = 0$. Possible ways in which this condition fails to hold (for example due to unobserved wealth), are extensively considered in the empirical analysis.

How well does e_I reflect the causal link between risk aversion and entry? For given r, the marginal effect of risk aversion on the entrepreneurship probability, i.e., the causal effect, is given by $dE(e|X,r)/dr$ from in (1). A more economically meaningful magnitude is the average marginal effect (averaging over r), which equals $E_r[dE(e|X,r)/dr]$. Since this magnitude is scale-dependent, let us focus on the average marginal effect of a one standard deviation increase in risk tolerance, i.e., $\sigma_r E_r[dE(e|X,r)/dr)]$ denoted by e_r. We now wish to link the causal magnitude e_r to the regression model (4). First define $r_I = [dE(r|X,I)/dI]/\sigma_r$. It follows directly from (2) and (3) that $r_I < 0$. Similarly to errors-in-variables models (e.g., Wooldridge, 2006), the following relation holds between e_I and e_r. The proof appears in Appendix A.

Remark 1

$$e_I = e_r r_I. \tag{5}$$

Thus, e_I is proportional to e_r. For stock market participation, the proportionality factor r_I captures the difference in mean risk aversion between investors and non-investors, expressed in standard deviation units. e_I will therefore be larger than e_r if the difference in mean risk aversion is larger than one standard deviation unit (i.e., $|r_I| > 1$) and lower if not.[11] Extant research from Germany and Norway suggest $|r_I| < 1$, so that estimates

[11] An example illustrates why a large $|r_I|$ leads to estimation difficulties. Suppose that I is a dummy that captures whether the individual is a parachut jumper or not. Parachuting is likely associated with low risk aversion (i.e., $|r_I| >> 1$). A large estimated e_I could therefore reflect just that parachut jumpers

of e_I will be conservative estimates of e_r.[12] We are not aware of evidence on r_I for the other risk tolerance proxies.

The variation in the proxy I in (4) conditional on controls is jointly determined by variation in risk tolerance and in additional (non-risk tolerance related) variation, captured by ϵ_I in (2) and (3). What happens if the proxy is weak, i.e., σ_I^2 is large? A larger σ_I^2 means that the proxy becomes more weakly related to risk tolerance and that $|r_I|$ tends to drop (see Appendix E for a formal analysis). From Remark 1 it follows that e_I tends towards zero as σ_I^2 becomes large. Similar to attenuation in measurement error models, therefore, a large σ_I^2 leads to e_I being a downward biased estimate of e_r.

Remark 1 relates closely to well-known properties of models with measurement error. For example, recall that in linear regression models with classic measurement error in a right hand side variable the coefficient on the mis-measured variable will be biased towards zero, due to attenuation (e.g., Wooldridge, 2006). We obtain this result in the special case $\delta_I = 1$ in (3).[13]

are much more risk tolerant than non parachut jumpers (i.e., a large $|r_I|$) rather than risk aversion playing a large role for the entry decision (i.e., $|e_r|$ being large).

[12] Dohmen et al. (2009) estimate individual risk tolerance based on data from an experimentally validated survey from Germany. The survey includes questions about stock ownership. In private communication, Uwe Sunde reports $|r_I|$ to be about 0.07. Aarbu & Schroyen (2011) estimate individual risk preferences using survey data from Norway. The survey includes the item "How likely is it that you would borrow money to invest in stocks?" In private communication, Karl Ove Aarbu reports $|r_I|$ to be about 0.30 (calculated based on a dummy variable that is one if an individual answered "somewhat likely", "likely", or "very likely" to the item).

These studies thus suggest self-selection of less risk-averse individuals into the stock market and, equally important, that stock market investors are not "pathological risk-seekers". In other words, $|r_I| < 1$ is a reasonable assumption to make, in which case e_I should provide us with conservative estimates of e_r.

The puzzlingly low estimated $|r_I|$ reported by Sunde is likely due to two factors. First, that "stock ownership" in their survey includes employee stock, inherited stocks, privately held stock, and mutual fund investments. Mutual fund investments are especially likely to be perceived less risky than common stock investments, and only moderately to involve self-selection on low risk-aversion. Second, the stock market participation rate increased considerably from 1995-1999 (our data period) to 2004 (Dohmen et al.'s data period) and it is plausible that the newcomers were on average more risk-averse than the existing investors.

[13] For $\delta_I = 1$ it can be shown that $|dE(r|X, I)/dI| < 1$ and by Remark 1 it follows that e_I will be a downward biased estimate of $E_r[dE(e|X, r)/dr)]$. To see the link to measurement error models more closely, consider the linear regression model $y = x\beta + \epsilon$. Suppose that $x^* = x + \eta$ is observable while x is not, and that x and η are independent. Then we have the well-known result (e.g., Bound et al., 2001) that $dE(y|x^*)/dx^* = \beta p < \beta$, because $p = \dfrac{\sigma_x^2}{\sigma_x^2 + \sigma_\epsilon^2} < 1$, where p is the precision ratio. To link to Remark 1, note that p is identical to the coefficient from a regression of x on x^* (the "inverse regression")

2.4 Performance

The entry model (1) does not specify firm performance as a separate entity. We now show that with a natural decomposition of (1), it follows that firms started up by risk tolerant entrepreneurs perform worse. Let firm profits (or, equivalently, entrepreneurial income) y be given by,

$$y = \alpha + X\beta_y + \epsilon_y, \tag{6}$$

where α is a random, fixed gain of entry (possibly negative), known to the individual, with density function $f_{\alpha,X}$. The term ϵ_y is random with $E(\epsilon_y|X,e,r) = 0$ and $Var(\epsilon_y|X) = \sigma_y^2$, whose realization is unknown at time of entry. For simplicity, we assume wage work offers constant income w so that $EU(w,r) = w$ (it is sufficient that wage work is less risky than entrepreneurship, see Appendix B). Suppose preferences are such that expected utility of an uncertain income stream with mean μ and variance σ^2 equals

$$EU(v,r) = \mu - r\sigma^2. \tag{7}$$

Therefore, an individual therefore becomes an entrepreneur if,

$$EU(y,r) = \alpha + X\beta_y - r\sigma_\tau^2 > w. \tag{8}$$

[To see that (1) is equivalent to (8) define $\alpha_E = \alpha - w$, $\beta_E = \beta_y$ and $\delta_E = \sigma_\tau$]. We then have the following (proof appears in Appendix B).

Remark 2 $dE(y|X,r,e=1)/dr > 0$.

The intuition behind this result is simple. If the marginal entrepreneur becomes more risk-averse, he must be compensated by a higher mean entrepreneurial return in order to still choose entrepreneurship. The same result holds in the Kanbur's (1979) model.[14]

---

i.e., $p = dE(x|x^*)/dx^* = \dfrac{cov(x,x^*)}{var(x^*)} = \dfrac{cov(x,x+\eta)}{var(x+\eta)} = \dfrac{\sigma_x^2}{\sigma_x^2+\sigma_\epsilon^2}$. In our setting, r plays the role of x, I plays the role of x^*, and $dE(r|X,I)/dI$ plays the role of p.

[14] In Kihlstrom & Laffont (1979), entrepreneurial income is assumed to be fixed so that $dE(y|X,r,e=1)/dr = 0$.

One type of entry model where Remark 2 would not necessarily hold is as follows. Suppose each individual selects an optimum entrepreneurial project from a menu of possible projects, and then selects whether to start up a firm or not. If the efficient menu of projects slopes upward in the $E(y)$, $Var(y)$ space, more risk tolerant entrepreneurs have optimum projects with a higher $Var(y)$ and higher $E(y)$. Such a model would have an ambiguous prediction for the relation between r and $E(y)$, due to the two counteracting effects. We show later that investors start up firms with statistically significant higher variance of operating returns on assets (OROA), compared to non-investors. However, the economic magnitude of the difference in variance seem small (Panels C and D of Appendix Table C4), so that the constant-variance assumption underlying Remark 2 appears to be a good approximation.

3 Data and summary statistics

The data comes from Norway and has been collected from several government registries. Socio-demographics, compiled by Statistics Norway, consist of yearly records of education, employment, location, income, and wealth information from 1993 to 2007. Norwegians are subject to a wealth tax, and submit a detailed annual overview of their assets to the tax authorities; the data are based on these reports. Earnings and wealth figures are public information in Norway. This transparency makes tax evasion more difficult and data more reliable. Statistics Norway data contains information on the population of Norwegian adults, not a sample as with the Panel Study of Income Dynamics or the Survey of Consumer Finance. Data on start-ups came from Bronnoysundregisteret, listing total equity, owners, and their ownership shares for the population of incorporated companies started in Norway between 2000 and 2007. It also contains an anonymous ID number for firms that could be matched with other data sources. Longitudinal accounting information was collected from Dun & Bradstreet, including annual data on firm performance such as sales, employment, and profitability. We follow each firm from the first year onward, and consequently there is no survivorship bias. Information on common stock transactions includes the population of trades for Norwegian individual investors, and is collected from the Norwegian Central Securities Depository. The dataset does not include information about mutual fund holdings.[15]

As in other industrialized countries, starting an incorporated company in Norway carries tax benefits relative to self-employment (e.g., more write-offs for expenses such as home office, company car, and computer equipment). With the exception of very small projects, incorporation is more tax efficient than self-employment status. The formal capital requirement for registering an incorporated company was NOK 100,000 (EUR 13,000) during the study period. Incorporated companies are required to have an external auditor certify annual accounting statements submitted to tax authorities.

We define an entrepreneur as a male with more than 50 percent ownership at incorporation in a firm started up in the period 2000 to 2007. To avoid counting wealth management vehicles as start-ups, we omit finance and real estate firms (NACE codes 65-70). The inclusion of these firms gives similar results. Restricting the definition of an entrepreneur to majority owners makes it unlikely to capture nominal founders such as "sleeping spouses". Restricting attention to males avoids measurement problems with female labor market participation.

To avoid endogeneity issues in the entry and performance regressions, we use predetermined values for control variables. The predetermined values are computed as averages across 1993 and 1994. The primary control variables are wage income, education in years, marital status, age, and household wealth. Wage income and education are likely strong correlates with cognitive skill and human capital. Wealth is likely to be correlated with the opportunity to explore both the stock market and entrepreneurship. In the analysis, we use several alternative measures and specifications for wealth. The richness of the data allows incorporation of a large number of additional control variables such as the numbers of children and siblings, region of residence, education type, and 1993 to 1994 employment characteristics (2-digit SIC codes and firm size). In additional specifications we control for family background, given as parent and spousal education and self-employment activity, parental wealth, and parental investor status. We do not have information about parents that are not part of the labor force, which reduces the sample size in these regressions.

[15] Doskeland & Hvide (2011) describe the stock market data and the Norwegian institutional environment in more detail, including questions of representativity. The data is anonymized, but contains an individual ID number that can be matched with the other data sources.

In the entrepreneurial performance regressions, we control for business cycle and firm age in addition to start-up size and industry (2-digit NACE codes). The remaining control variables are similar to the entry regressions.

To eliminate individuals under training or close to retirement, we restrict the sample to individuals between 25 and 50 years old in 1993. The youngest founder is 32 and the oldest 64 at the incorporation date. We eliminate individuals who are unemployed or self-employed in 1993 or 1994. We also eliminate individuals that work for a listed company or a subsidiary between 1993 and 1999, as they are likely to receive company stock as part of their compensation package; a small remuneration in stock each year received favorable tax treatment relative to cash payment. We also eliminate a small fraction of individuals with missing values for one or more control variables. Our sample of males with complete 1993-1994 socio-demographic records is about 397,019 individuals, of which about 6,307 become entrepreneurs (about 1,6 percent). For individuals who start up more than one firm in the database, we choose the first entrepreneurial spell for the performance analysis. Specifics for the sample of entrepreneurs – in terms of entry by year, industry, and inactivity by year – are described in detail in Appendix C, Table C1. The timeline for the data is described in Appendix C, Figure C1.

Table 1 *here*

Table 1, Panels A and C show descriptive statistics for the individuals and start-ups in the sample. In Panel A (column 1), we report descriptive statistics for the full sample of investors (column 2) and non-investors (column 3).[16] Panel C, column 12 reports start-up descriptives, and columns 13 and 14 distinguish between investors and non-investors respectively. In Panel B, columns 8 through 11, we report average stock market activity from 1995 to 1999 (column 8), and differences for entrepreneurs and non-entrepreneurs

[16] The comparison between entrepreneurs and non-entrepreneurs is not presented due to space considerations. Consistent with previous studies, entrepreneurs are on average wealthier, have a higher income, and are more likely to be married than non-entrepreneurs. The means are close to those reported in previous studies using U.S. data (Hamilton 2000; Hurst and Lusardi 2004; Campbell 2006). Moreover, the start-ups are small; on average, they possessed NOK 4.5 million in assets in the first year (EUR 550,000), with a much lower median. The average start-up size is similar to that in the 2003 Survey of Small Business Finance from the United States.

(columns 9 and 10). All NOK figures are inflation-adjusted, using 2002 as the base year. Norway has a relatively high stock market participation by international standards; about 17 percent of the subjects own common stocks sometime between 1995 and 1999. This is similar in the U.S. for the same period (Bogan, 2008). Investors have a considerably higher probability of becoming entrepreneur than the non-investors (2.88 percent for investors and 1.32 percent for the non-investors).[17] We can also note from Panel C that companies started up by investors are less profitable in terms of operating return on assets (OROA), have fewer employees and a lower probability of 4-year survival than companies started up by the non-investors.

4 Entry

We now explore the relationship between revealed risk preferences and the propensity to start up a firm, based on (4). The dependent variable is a dummy that equals 1 if an individual starts up a firm in the period 2000 to 2007, and 0 if not. The main independent variables are proxies for risk tolerance based on past investment behavior. Across all regressions, we control for human capital level with age (second-order polynomial), years of education and log(income). We control for access to liquidity by using log(gross household wealth).[18] We control for type of human capital by type of education dummy variables, employer size and industry, and a dummy variable representing whether the individual was self-employed sometime between 1986 and 1993. Dummy variables for region of residence and logarithms for numbers of children and siblings are also incorporated.

In Table 2, we report coefficient estimates based on a linear probability model. To

[17] Table 1, columns 5 and 6 compare investors with a matched sample of non-investors. The matching procedure, described in Appendix D, ensures that investors and non-investors are similar with respect to socio-demographic characteristics. Note that in the matched sample, investors have a significantly higher probability of entry (2.88 percent), compared to non-investors (2.05 percent).

[18] Taxable wealth is a noisy measure of true wealth because the value of property investments and investments in non-listed stocks has an artificially low tax value. The tax value of real estate is a maximum of 30% of market value; non-listed stock is valued at book value. The tax value of debt is close to market values. Financing property and non-listed stocks with debt is a common way to avoid wealth taxation. For this reason, gross taxable wealth is likely at least as good a proxy for true wealth as net taxable wealth. We test results for robustness different measures of wealth.

accommodate non-linearities, we also perform probit regressions.[19]

Table 2 here

In Table 2, Panel A, columns 1 and 2 we estimate e_I controlling for wealth using a first-order and a second-order polynomial of log(household wealth). The estimated e_I in column 2 is about 0.82 percent. Since the baseline probability of becoming an entrepreneur is about 1.59, being an investor increases the entry probability by about 50 percent. The estimate of e_I in columns 1 and 2 would be upward biased if investor status captured higher unobserved wealth. In column 3, our main specification, we use a fifth-order polynomial for wealth, and include a third-order polynomial in salary income and an interaction term between income and wealth (as in Hurst & Lusardi, 2004, a fifth-order polynomial for wealth fits the data better than including a quadratic in wealth because the data includes a non-negligible fraction of individuals with very large or very small wealth values). The estimated e_I in column 3 is of very similar magnitude as in column 2, which does not support that investor status captures unobserved wealth.[20]

Table 2, column 4 and 5 confine attention to the subsample of individuals for whom we know parental wealth, parental and spousal self-employment, and parental business education status. Parental self-employment and investor status serve as a proxy for a business-oriented, entrepreneurial, family background. In column 4, the estimated e_I for this subsample is larger than in column 3 (59 percent), and in column 5 the estimated effect is very similar to that of columns 2 and 3. If risk preferences are correlated inside the family, as suggested by the work of Dohmen et al. (2011), parental investor status would partially capture own risk tolerance, which could explain why the estimated effect is smaller in column 5 than in column 4.

To investigate interactions, we estimate e_I for various subgroups using the same set of controls as in column 3. These results are reported in Appendix C, Table C3. Across a

[19] Non-linearities occur because e_I gets smaller as predicted entry rate gets close to 0 or 1. Observe from (4) that $e_I = \gamma f(\alpha + X\beta + I\gamma)$, where $f()$ is the density function of ϵ. This implies that e_I is non-linear except when $f()$ is a constant.

[20] The predicted relation between wealth and entry is non-linear and quite similar to in Hurst & Lusardi (2004). See Figure C2 in Appendix C for two plots that illustrate the relation between wealth and entry.

wide array of subgroups (i.e. including individuals that live outside the Oslo region, the main metropolitan area in Norway, that have below median wealth, or excluding those with business education, the previously self-employed, and excluding family firms, among others) the estimated e_I is statistically and economically significant. For most subgroups, the estimated e_I is between 35 and 50 percent. One interesting finding is that e_I is larger for young individuals: the estimated effect for age group 25 to 33 is around 74 percent. The estimated effect is also larger for single individuals and individuals in the lowest wealth quartile.

To accommodate that individuals with low wealth or income are unlikely to invest in the stock market or start a firm, we also report the results from weighted regressions (OLS and probit). The regression weights are calculated using two nearest-neighbor propensity score matching (the outcome variable in the first stage regression is a stock ownership dummy variable and the matching procedure is described in detail in Appendix D). Appendix C, Table C2 reports the main entry results using three alternative regression models: a linear probability model with 2 nearest-neighbor propensity score matching (Panel A), a probit model without matching (Panel B), and a probit model with 2 nearest-neighbor propensity score matching (Panel C). The magnitudes of the risk tolerance proxies are smaller than those reported in Table 2 but are still large and significant.[21]

Panel B considers the relation between entrepreneurial entry and other proxies for risk tolerance.[22] As described in Section 2, asset allocation theory suggests that the fraction of wealth invested in risky assets should be proportional to risk tolerance. In columns 10 and 11 we report results using log(fraction of wealth invested in stock market), averaged across 1995 to 1999, as a proxy for risk tolerance, and in columns 12 and 13 we report results using log(fraction of income invested in the stock market). The results are similar not using logs. The log specification makes our estimates less sensitive to outliers. In columns 11 and 13 we include only investors, and in columns 10 and 12 we use the full

[21] We define a stock market investor as somebody with a positive holding of stocks during 1994-1999. We can define investor status more stringently, by individuals that purchase stocks during this period. The results from these estimations are shown in the lines 26 and 27 of Appendix Table C3. The estimates show that investors, defined in this fashion, are 65 and 68 percent more likely to become entrepreneurs. Our results on entrepreneurial performance, reported in the next section, are also robust.

[22] The correlation matrix for the proxies is reported in the Appendix Table C6.

sample by setting portfolio value to zero for the non-investors (and including a dummy variable for non-investor status). Across the four columns, there is a strong, positive relationship between fraction of wealth and income invested in the stock market and predicted probabilities of entry. An increase in portfolio value to wealth ratio from the 25th percentile to the 25th percentile (hereafter interquartile increase), i.e. from 0.002 to 0.104, increases predicted entry by one third. The effect remains the same when using the log of portfolio value to income ratio as a risk proxy. The effects are of the same magnitude to the estimated e_I in Panel A, in the estimates for the pooled sample controlling for non-investor status (columns 10 and 12). In columns 14 and 15 we proxy risk tolerance with debt-to-wealth and debt-to-income ratios. Also leverage ratios correlate positively with entry. An interquartile increase in the debt to income ratio, from 0.64 to 2.15, increases the likelihood of entry by 11.5 percent. In unreported regressions, we show that the coefficient estimates are slightly reduced if investor status is included as a control. As a fourth proxy for risk tolerance, we use the volatility of portfolio returns, measured by the standard deviation of monthly portfolio returns, controling for portfolio value with a second-order polynomial. This proxy for risk tolerance is more appropriate than market (beta) risk because most individual investors have undiversified portfolios. In unreported regressions, we show that using volatility of portfolio returns as proxy for risk tolerance gives very similar results to those in columns 10 to 15. We conclude that across a range of complementary proxies for risk tolerance, there is a strong positive association between risk tolerance and entrepreneurial entry.

Since personal wealth affects both the stock market participation and the entrepreneurship decision, we investigate the possibility of omitted variable bias due to unobserved wealth. We investigate this question by using several alternative measures of wealth, in Panel C. Column 16 uses the same specification as column 3 except a fifth-order polynomial in log individual wealth rather than log household wealth. Column 17 uses net household wealth (fifth-order polynomial) rather than log gross household wealth as a control. In column 18, we use a "flow", net capital income (fifth-order polynomial), rather than the "stock" log(gross wealth). The estimated e_I changes only marginally when using log individual wealth in column 16. The magnitude of the effect increases to 57 and 59 percent, in the specifications with net capital income and net wealth, respectively. Investment in the stock market may have a positive average liquidity effect relative to safer investments, leading to less binding wealth constraints and more entry. Therefore, we re-estimated (4) excluding stock owners whose portfolio returns exceeded the returns on government bonds in the period 1995 to 1999. The estimated coefficient for individuals whose realized liquidity effect of stock ownership is negative relative to safe investments is reported in line (27) of Appendix C Table C3. The effects are larger than in the full sample. To capture liquidity effects alternatively, we also estimate e_I using 1999 measures of wealth, replacing the 1993-94 wealth measures. The estimated coefficient is large but smaller than in Panel A, around 32 percent, perhaps because of the endogeneity of 1999 wealth with respect to the entrepreneurship decision. This result is also reported in Appendix C Table C3 (line 33). Finally, we incorporate a dummy variable capturing home ownership (defined as real estate ownership greater than NOK 100,000) in the specification of Table 2, Column 3. The result of this exercise is reported in line (29) of Appendix C Table C3, showing that the effect of investment on entry is robust. All the remaining results in this paper are robust to the inclusion of the home ownership variable. Overall, these robustness exercises suggest that the results reported in Panel A and B are not driven by unobserved liquidity effects.

We do not observe entrepreneurial entry before 2000, and may underestimate the relationship between risk tolerance and entrepreneurship if the most risk tolerant start a company before 2000. One way to account for this possible bias is to use the case control-matching methodology, commonly used in epidemiology.[23] Column 9 of Table 2 reports conditional logit estimates using case control matching in which every entrepreneur is matched with twenty non-entrepreneurs based on age, education, marital status, wealth, and salary groups. The estimated odds ratio is 1.5, meaning that investors are 1.5 times more likely to become entrepreneurs than non-investors. This is within the same order of magnitude as in column 3, and suggests that bias due to "early entry" is not large.

Another concern is that many of the start-ups in our sample could be wealth management vehicles, or an advanced form for leisure activity. We already dropped real estate and financial companies. An additional way to capture "real" start-ups is to confine at-

[23] This methodology has, among others, been used to infer a relation between smoking and lung cancer with "backward-looking" samples.

tention to hi-tech startups. In columns 7 and 8 of Table 2, we examine this possibility using a narrow and a wide definition of hi-tech firms.[24] In column 7 and column 8, we find that investor status explains about 70 percent of entry rates. In unreported analysis, we define entrepreneurship as starting up a company that has above NOK 300,000 initial equity. The estimated percentage effect of being an investor is very similar to in the Table 2. Thus our results are not driven by most start-ups in the sample being very small.[25]

The results in Table 2 support Knight's (1921) conjecture that more risk tolerant individuals self-select into entrepreneurial activity. That the association between entry and stock market participation is statistically significant is in itself not particularly surprising. What we find remarkable is the consistency of the effects across different proxies for risk tolerance, and the magnitudes of the estimated effects. The results suggest that risk tolerance plays a large role for entrepreneurial entry. In the next section we consider the additional prediction, formalized in Remark 2, that more risk tolerant individuals start up firms of inferior quality.

5 Firm performance

We now examine whether firm performance is negatively associated with entrepreneurial risk tolerance, as posited by Remark 2. The performance analysis is based on a yearly accounting panel for the period 2000 to 2010. The panel tracks every firm in the sample from its first year onwards until its death (if applicable). Hence the oldest firms in our sample have been in operation for ten years. As performance measures, we use number of employees to measure job creation, sales to measure growth, and OROA to measure profitability.[26] In addition, we use 4-year business survival. The main indepen-

[24]By hi-tech firms we mean firms started up in hi-tech sectors. We define hi-tech sectors narrowly as the following NACE codes: 24420; 30; 32; 33; 64; 72; 73. Medium-high-tech sectors, our broader definition, also include the following the NACE codes: 24; 29; 31; 34; 35; 61; 62; 64; 65; 66; 67; 70; 71; 72; 73; 74; 80; 85; 92).

[25]Column 6 of Table 2 reports the estimated e_I when the dependent variable is self-employment entry between 1995 and 2007. Self-employment is a more noisy measure of entrepreneurship than starting up an incorporated company (for example, self-employment is more likely to be hidden unemployment) and the estimated coefficient is as expected smaller.

[26]OROA is the ratio of earnings before interest and taxes (EBIT) to the total asset base used to generate them, the standard performance measure in the accounting and financial economics literature

dent variables are two proxies for risk tolerance; stock market participation and personal leverage.[27] We account for firm age and industry effects by using yearly dummies for firm age and dummies for start-up industry groups (2-digit NACE codes), and for business cycle effects by using year dummies. In addition we control for the same pre-determined socio-demographic variables as in the entry analysis (excluding 2-digit industry codes for the place of employment in 1993, as we already include dummies for startup industry groups), and for start-up size measured by a second order polynomial in log(first year equity).

As a preliminary step, we investigate whether firms started up by the more risk tolerant are different with respect to size or financing. We regress start-up size at the incorporation date, measured by log(equity), on the risk tolerance proxies, using the same socio-demographic controls as in Table 2, adding year of incorporation dummy variables to the set of controls. The analysis (results reported in Table C4, Panel A) suggest that firms started by more risk tolerant individuals are not larger than firms started by less risk tolerant individuals. Thus a higher risk tolerance is associated with a higher rate of entrepreneurial entry, but not with larger start-ups. We next investigate whether firms started up by the more risk tolerant are financed differently. The analysis (reported in Table C4, Panel B) suggests that, after controling for size measured by a second order polynomial in log(equity), firms started up by more risk tolerant individuals have a higher leverage ratio in their first year. The effects are rather small; the estimated elasticity of firm leverage ratio to personal leverage ratio is about 3 percent.

We also investigate whether firms started up by investors have a more volatile return, as measured by the standard deviation of OROA (recall that in the model underlying Remark 2, we assumed constant variance across projects). The analysis suggests that, after controling for size measured by a second order polynomial in log(equity) and for

(Bennedsen et al. 2007 and references therein). Unlike returns to equity or returns to capital employed, OROA compares firm profitability relative to total assets. In contrast to net income measures such as return on assets, OROA is unaffected by capital structure or dividend policy differences across firms.

[27] In unreported analysis, we regress performance on the portfolio-based measures of risk tolerance (portfolio value to wealth and income, and volatility of monthly portfolio returns). The sample size is severely reduced because we confine attention to individuals that are both investors and start up a firm. The coefficients on the risk tolerance proxies we obtain are negative and economically large for survival but statistically insignificant. For the other performance measures, the coefficients are neither economically nor statistically significant.

2-digit startup industry codes (Appendix Table C4, Panel C), firms started up by more risk tolerant individuals have a more volatile OROA. In Panel C, a one percent increase in personal leverage ratio increases the standard deviation of OROA by 0.0003 (an interquartile increase in debt-to-income ratio increases the standard deviation by 1.6 percent) and firms started up by investors have a higher standard deviation of OROA by 0.009. These effects are economically of very small magnitude. There is substantial heterogeneity in the risk of a new business, and it is surprising that the firms started by the risk tolerant investors are no more risky. One reason why Panel C of Table C4 may not adequately capture differences in risk is that we include industry dummies; thus we do not capture the possibility that more risk tolerant entrepreneurs enter industries with higher risk. One way to capture this possibility is to analyze whether risk tolerant entrepreneurs are more likely to enter hi-tech industries. As reported in Table 2, column (7), this is indeed the case; we find a 74 percent investor effect in column (7), the main regression, and a 70 percent investor effect in column (8). Another way to capture the possibility of selection into more risky industries by more risk tolerant individuals is to drop industry dummies from the volatility regressions in Table C4, Panel C. We find that the investor effect increases from 4.6 percent in Panel C to 5.7 percent in Panel D, where we have dropped industry dummies in the latter. The difference appears too small to grossly invalidate the constant variance assumption underlying Remark 2.

The following table reports results from panel regressions where measures of entrepreneurial performance are regressed on investor status and personal leverage.

Table 3 *here*

Table 3, Panel A shows all performance measures associate negatively with stock market participation. For example, start-ups in which the entrepreneur was a stock market investor have about 0.012 lower yearly OROA (corresponding to an effect of 12.7 percent, given the average OROA in the sample). They have 22 percent lower sales and 27 percent lower employment. They have about 2.55 percentage points lower probability of surviving 4 years (corresponding to a 4.3 percent effect). Panel B reports results using personal leverage as a proxy for risk tolerance. Personal leverage is negatively associated with both profitability, growth, and survival. An interquartile increase in the leverage ratio results in about 8 percent reduction in OROA, 6 percent reduction in employment and a 3 percent reduction in sales and 4-year survival.[28] In unreported regressions we use debt to wealth rather than debt to income as a measure of personal leverage, obtaining very similar results.

The results in Section 2 are consistent with learning effects from stock market investment (or correlates) that increases entrepreneurial productivity and entry rates.[29] Less so for the results of the current section; it is unlikely that learning leads to poorer firm performance. Taken together, the results in Table 2 and Table 3 therefore give support to the self-selection mechanism highlighted in Remark 2: more risk tolerant individuals are more inclined to start up a firm but of poorer quality. We cannot rule out that more risk tolerant individuals perform worse partially because they run their firms differently (it is, however, unlikely that risk tolerance is associated with an excessive initial investment as we find no difference in initial firm size in Table C4). Neither can we rule out that risk tolerance is correlated with a high cost of effort that leads to lower effort and an inferior entrepreneurial performance. This seems unlikely, however, since we include prior wages, age, and education, in our set of controls, and these variables are likely to at least partially capture differences in the cost of effort. One explanation of our findings is that overconfidence is correlated both with the risk tolerance proxies and with entrepreneurial entry and performance. This alternative mechanism is considered in the next section.

6 Behavioral Effects

The empirical literature suggests that behavioral effects affect asset allocation (see e.g., Campbell, 2006, for a review). For example, individual investors tend to own too few stocks to be well-diversified, and trade too much for their own good (e.g., Odean, 1999;

[28]In the panel sample, the 25th percentile has 0.99 leverage ratio, and the 75th percentile has 2.62 leverage ratio.

[29]The correlation between stock market investments and entry could be driven by unobserved IQ. We have access to IQ data for a subsample of the entire sample. In unreported regressions, we find that IQ has explanatory power for stock market participation (see Grinblatt et al., 2011, for a similar result), but no explanatory power on entrepreneurial entry, or on performance. A likely reason is that the previous wage and wealth controls absorbs the influence of IQ on entrepreneurship.

Barber & Odean, 2000). Odean (1998) and Barber & Odean (2000) suggest excess trading is due to overconfidence, a trait that has been suggested as associated with entrepreneurial entry (e.g., Landier & Thesmar, 2007).

In the following we examine whether controlling for overconfidence affects the estimated relationship between risk tolerance and entrepreneurship (in unreported regressions we do not find any association between overconfidence and start-up size). We use trading intensity as a proxy for overconfidence. We also attempt to control for optimism by using a proxy derived from Puri & Robinson (2007); owning one stock.[30] The proxies are defined only for stock market investors, therefore the sample size is smaller than in Table 3.

Table 4 here

In Table 4, Panel A, column 1, we find a strong, positive relationship between entrepreneurial entry and the proxy for overconfidence, log(number of trades). Sales growth is associated negatively with log(number of trades), while for the other performance measures there is no relationship. In Panel B, there is a statistically significant negative relationship between entrepreneurial entry and our proxy for optimism, owning one stock. The proxy correlates positively with sales. In Panel C we find a strong, positive relationship between entrepreneurial entry and personal leverage, and a negative relationship between leverage and performance, confirming previous estimates for the full sample (which includes also non-investors). Finally, in Panel D we regress entry and performance on the risk tolerance, overconfidence, and optimism proxies. Overconfidence seems to play an important role in entrepreneurial entry but no role for performance. Moreover, it seems unlikely that the positive relation between risk tolerance and entrepreneurship found in Tables 2 and 3 are driven by unobserved overconfidence or optimism. The reason is that including controls for these effects has little or no impact on the relationships between risk tolerance and entrepreneurial entry and performance (compare Panel C and Panel D); the role of overconfidence seems largely orthogonal to the role played by risk tolerance.

In Table C5, Appendix C we attempt to accommodate the role of sensation-seeking

[30] Puri & Robinson (2007) define optimism as having a subjective life expectancy that is higher than actual life expectancy. They find a correlation between owning one stock and being an optimist.

preferences (Grinblatt & Keloharju, 2009) by using individual data on car ownership from 1999.[31] We incorporate log(car horse power/size), where size is measured as car length multiplied by width. Column 1 of Panel A implies that an interquartile increase in horse power/size increases the entry probability by 16 percent. Panel B shows that stock market participation exerts a similar effect on entry in the car owner sub-sample as in Table 2. However, its effects on firm performance are larger than in Table 3. In the car owner sub-sample, firms initiated by investors have somewhat 48 percent lower OROA, 43 percent lower sales, 35 percent fewer employees and 19 percent lower survival rates. Panel C shows that when stock ownership is incorporated jointly with the horse power variable both effects remain significant and at the magnitudes estimated in the previous two panels. Investors are about 37 percent more likely to become entrepreneurs. The results using leverage as a risk tolerance proxy are similar and not reported.

In conclusion, although we find support for behavioral effects affecting entrepreneurial entry, we do not find support for behavioral effects driving the documented relationship between risk tolerance and entrepreneurship; all our main results are robust to including controls for behavioral effects.

7 Conclusion

In this paper, we explore the origin of firms by focusing on the founders. We find evidence in favor of the Knight (1921) hypothesis that more risk tolerant individuals are more likely to start up firms. We use several proxies to capture revealed risk preference: stock market participation, personal leverage, and fraction of wealth invested in the stock market. In addition, we find evidence that firms started up by more risk tolerant entrepreneurs perform worse. All our results are consistent with a simple self-selection story: more risk tolerant individuals are more inclined to start up a firm but at the margin start up firms of poorer expected quality than less risk tolerant individuals.

These results point to the selection of individuals based on differences in risk preferences as an important element in the origin and growth of firms. Much of the existing

[31] The car ownership data is incomplete and renders 26,665 matches with the full sample. The individuals covered by the car data are younger and somewhat wealthier.

evidence in favor of the risk tolerance hypothesis comes from comparing the variability of returns for entrepreneurs and wage workers (e.g., Heaton and Lucas, 2000, Hamilton, 2000, Hall & Woodward, 2010). However these findings are open to several interpretations (e.g., Vereshchagina & Hopenhayn, 2009). A key contribution of our analysis is to directly measure the individuals that are more risk tolerant, and to show that these individuals are more likely to start up firms. This field evidence complements the vast experimental and psychological evidence on individual risk tolerance.

Our results also have important implications for models of industry evolution. We find that the performance of entering firms is correlated with the founder's risk attitude. This stands in contrast to standard models of entry where firms are assumed to be ex-ante homogenous, and suggests the relevance of theories where entrepreneurial heterogeneity is taken into account. Also, our findings suggest that individual entrepreneurs may account for some of the unexplained differences in performance for young firms. For example, one reason why many young businesses fail in the beginning may be because of self-selection of risk tolerant (or even risk-loving) individuals with relatively poor ideas into entrepreneurship.

We highlight two areas of future research. Our results suggest that cross-sectional variation in risk aversion can explain a substantial amount of cross-sectional variation in start-up activity. It would be of interest to examine whether time-series variation in risk aversion can explain time-series variation in start-up activity, where time-series variation in risk aversion can occur due to evolution in risk preferences at the individual level (through e.g., wealth shocks or business cycle variations, as in Rampini, 2004) or due to cohort effects (as in Malmendier & Nagel, 2011). Possibly, such an extension could lead us one step further in understanding why start-up activity varies so much over time. A second extension would be to investigate whether differences in entrepreneurship rates across countries are attributable to differences in the risk-preference distributions. For example, with a small sample of experimental subjects, Weber & Hsee (1998) find that Chinese subjects place higher value on risky financial options than German and U.S. subjects. This finding could possibly link to the seemingly high entrepreneurship rates in China (Djankov et al., 2006).

The general message is that examining individual heterogeneity in more detail can lead to a better understanding of the origin and growth of firms.

8 References

Aarbu, K. O. and F. Schroyen (2011). Mapping risk aversion in Norway using hypothetical income gambles. Working paper, Norwegian School of Economics and Business.

Andersen, S., G. W. Harrison, and E. Rutstrøm (2008). Lost in state space: are preferences stable. International Economic Review, 49, 1091-1112.

Andersen, S. and K. M. Nielsen (2012). Ability or finances as constraints on entrepreneurship? Evidence from survival rates in a natural experiment. Review of Financial Studies, 25, 3684-3710.

Angrist, J. and J.-S. Pischke (2009). Mostly harmless econometrics. Princeton, NJ. Princeton University Press.

Asplund, M. and V. Nocke (2006). Firm turnover in imperfectly competitive markets. Review of Economic Studies, 73, 295–327.

Audretsch, D. B. and T. Mahmood (1995). New firm survival: new results using a hazard function. Review of Economics and Statistics, 77, 97-103.

Barber, B. and T. Odean (2000). Trading is hazardous to your wealth: the common stock investment performance of individual investors. Journal of Finance, 55, 773-806.

Barseghyan, L., F. Molinari, T. O'Donoghue, and J. Teitelbaum (2011) The nature of risk preferences: evidence from insurance choices. American Economic Review, 101, 591–631.

Barsky, R. B., F. T. Juster, M. S. Kimball, and M. D. Shapiro (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. Quarterly Journal of Economics, 112, 537-579.

Bennedsen, M., K. M. Nielsen, F. Pérez-González, and D. Wolfenzon (2007). Inside the family firm: the role of families in succession decisions and performance. Quarterly Journal of Economics, 122, 647-691.

Bertrand, M. and A. Schoar (2003). Managing with style: The effect of managers on firm policies. Quarterly Journal of Economics, 118, 1169-1208.

Bitler, M. P., T. J. Moskowitz, and A. Vissing-Jørgensen (2005). Testing agency theory with entrepreneur effort and wealth. Journal of Finance, 60, 539-576.

Bogan, V. (2008). Stock market participation and the internet. Journal of Financial and Quantitative Analysis, 43, 191-212.

Bound, J., C. Brown, and N Mathiowetz (2001). Measurement error in survey data. Handbook of econometrics. Elsevier: Chapter 59, 5, 3705-3843.

Buera, F. (2008). Persistency of poverty, financial frictions, and entrepreneurship. Mimeo, Northwestern University.

Caggese, A. (2012). Entrepreneurial Risk, Investment and Innovation. Journal of Financial Economics. 106, 287-307.

Campbell, J. Y. (2006). Household finance. Journal of Finance, 61, 1553-1604.

Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. Journal of Finance, 56, 1-43.

Chalice, D. R. (2001). How to differentiate and integrate sequences. The American Mathematical Monthly, 911-921.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007). Consistency and heterogeneity of individual behavior under uncertainty. American Economic Review, 97, 1921-1938.

Cocco, J., F. J. Gomes, and P. J. Maenhout (2005). Consumption and portfolio choice over the life-cycle. Review of Financial Studies 18, 491-533.

Cramer, J. S, J. Hartog, N. Jonker, and C. M. Van Praag (2002). Low risk aversion encourages the choice for entrepreneurship: an empirical test of a truism. Journal of Economic Behavior and Organization, 48, 29-36.

Djankov, S., Y. Qian, G. Roland, and E. Zhuravaskaya (2006). Who are China's entrepreneurs? American Economic Review Papers and Proceedings, 96, 348-362.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner. (2009). Individual risk attitudes: measurement, determinants and behavioral consequences. Journal of the European Economic Association. 5517, 522-550.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2011). The intergenerational transmission of risk and trust attitudes. Review of Economic Studies. 79, 1-33.

Doskeland, T. and H. K. Hvide (2011). Do individual investors have asymmetric information based on work experience? Journal of Finance. 66, 1011-1041.

Edgerton, D. and P. Jochumzen (2003). Estimation in binary choice models with measurement errors. Lund University, Department of Economics WP 2003, 4.

Evans, D. S. and B. Jovanovic (1989). An estimated model of entrepreneurial choice under liquidity constraints. Journal of Political Economy. 97, 808-827.

Gompers, P. A., J. Lerner, and D. Scharfstein (2005). Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986 to 1999. Journal of Finance, 60, 577-614.

Grinblatt, M. and M. Keloharju (2009). Sensation seeking, overconfidence, and trading activity. Journal of Finance, 64, 549-578.

Grinblatt, M. M. Keloharju, and J. Linnainmaa (2011). IQ and stock market participation. Journal of Finance, 66, 2121-2164.

Guiso, L. and T. Japelli (2005). Awareness and stock market participation. Review of Finance, 9, 537-567.

Guiso, L., P. Sapienza, and L. Zingales (2008). Trusting the stock market. Journal of Finance, 63, 2557-2600.

Hall, R. E. and S. Woodward (2010). The burden of nondiversifiable risk of entrepreneurship. American Economic Review, 100, 1163-1194.

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

Heaton, J. and D. Lucas (2000). Portfolio choice and asset prices: The importance of entrepreneurial risk. Journal of Finance, 55, 1163-1198.

Hong, H., J. D. Kubik, and J. C. Stein (2004). Social interaction and stock market participation. Journal of Finance, 59, 137-163.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. Econometrica, 60, 1127-1150.

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

Hvide, H. K. (2009). The quality of entrepreneurs. Economic Journal, 119, 1010-1035.

Hvide, H. K. and J. Moen (2010). Lean and hungry or fat and content? Entrepreneurs' wealth and start-up performance. Management Science, 56, 1242-1258.

Jones, B. F. and B. A. Olken (2005). Do leaders matter? National leadership and growth since World War II. Quarterly Journal of Economics, 120, 835-864.

Jovanovic, B. (1982). Selection and the evolution of industry. Econometrica, 50, 649-670.

Kanbur, S. M. (1979). Of risk taking and the personal distribution of income. Journal of Political Economy, 87, 769-797.

Kaplan, S. N., M. M. Klebanov, and M. Sorensen (2012). Which CEO characteristics and abilities matter? Journal of Finance, 67, 973-1007.

Kerr, W. R., J. Lerner, and A. Schoar (2011). The consequences of entrepreneurial finance: a regression discontinuity analysis: Evidence from angel financings. Review of Finance. Fortcoming.

Kihlstrom, R. E. and J. J. Laffont (1979). A general equilibrium entrepreneurial theory of firm formation based on risk aversion. Journal of Political Economy, 82, 719-748.

Klepper, S. and E. E. Leamer (1984). Consistent sets of estimates for regressions with errors in all variables. Econometrica, 52, 163-183.

Knight, F. (1921). Risk, uncertainty and profit. Boston, MA: Houghton Mifflin Co.

Kumar, A. (2009). Who gambles in the stock market? Journal of Finance, 64, 1889-1993.

Landier, A. and F. Thesmar (2009). Financial contracting with optimistic entrepreneurs. Review of Financial Studies. 22, 117-150.

Lazear, E. (2005). Entrepreneurship. Journal of Labor Economics, 23, 649-680.

Lerner, J. and U. Malmendier (2011). With a little help from my (random) friends: success and failure in post-business school entrepreneurship. NBER Working paper no. 16918. Forthcoming, Review of Financial Studies.

Leuven E. and B. Sianesi. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. http://ideas.repec.org/c/boc/bocode/s432001.html

Lucas, R. E. (1978). On the size distribution of business firms. The Bell Journal of Economics, 9, 508–523.

Malmendier, U. and S. Nagel (2011). Depression babies: Do macroeconomic experiences affect risk-taking? Quarterly Journal of Economics, 126, 373-416.

Malmendier, U. and G. Tate (2008). Who makes acquisitions? CEO overconfidence and the market's reaction. Journal of Financial Economics, 89, 20-43.

Merton, C. (1969). Lifetime portfolio selection uncertainty: The continuous-time case. Review of Economics and Statistics, 51, 247-257.

Moskowitz, T. and A. Vissing-Jorgensen (2002). The returns to entrepreneurial investment: A private equity premium puzzle? American Economic Review, 92, 745-778.

Mossin, J. (1968). Optimal multiperiod portfolio policies. Journal of Business, 41, 215-229.

Nanda, R. and J. Sorensen (2010). Workplace peer effects and entrepreneurship, Management Science, 56, 1116–1126.

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. Journal of Finance, 53, 1887-1934.

Odean, T. (1999). Do investors trade too much? American Economic Review, 89, 1279-1298.

Parker, S. (2009). The economics of self-employment and entrepreneurship. Cambridge University Press.

Puri, M. and D. Robinson (2007). Optimism and economic choice. Journal of Financial Economics, 86, 71-99.

Rajan, R. (2012). Presidential address: The corporation in finance. Journal of Finance, 67, 1173–1217.

Rampini, A. A. (2004). Entrepreneurial activity, risk, and the business cycle. Journal of Monetary Economics, 51, 555-573.

Robb, A. and D. Robinson (2012). The capital structure decisions of new firms. Review of Financial Studies, Forthcoming.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70, 41-55.

Sahm, C. (2007). How much does risk tolerance change? Working paper 2007-66, Federal Reserve Washington.

Samuelson, P. (1969). Lifetime portfolio selection by dynamic stochastic programming. Review of Economics and Statistics, 51, 239-246.

Shane, S., E. A. Locke, and C. J. Collins (2003). Entrepreneurial motivation. Human

Resource Management Review, 13, 257-279.

Syverson, C. (2011). What determines productivity? Journal of Economic Literature, 49, 326–365.

Vereshchagina, G. and H. A. Hopenhayn (2009). Risk taking by entrepreneurs. American Economic Review, 99, 1808-1830.

Vissing-Jorgensen, A. (2002). Towards an explanation of household portfolio choice heterogeneity: nonfinancial income and participation cost structure. Working paper, Northwestern University.

Weber, E. U. and C. K. Hsee (1998). Cross-cultural differences in risk perception but cross-cultural similarities in attitudes towards risk. Management Science, 44, 1205-1217.

Wooldridge, J. M. (2006). Introductory econometrics: a modern approach. Mason, OH: Thomson/South-Western.

Wu, B. and A. M. Knott (2006). Entrepreneurial risk and market entry. Management Science, 52, 1315-1330.

9 Appendix A

This appendix proves Remark 1 and Remark 2.

9.1 Proof of Remark 1

We start out with the case where I equation is linear, i.e., (3), and then consider the case where it is non-linear, i.e., (2).[32] Throughout the proof $E(.)$ denotes the expectation over ϵ_E while other expectation functions are subscripted. For convenience, we let $\alpha_E = 0$. The proof is straightforward to generalize to the case where α_E is non-zero.

Step 0. Remark 1 states that $e_I = e_r r_I$, which is equivalent to

$$dE(e|X, I)/dI = E_r[dE(e|X, r)/dr] \cdot dE(r|X, I)/dI. \tag{9}$$

We prove (9) in the following.

[32] Klepper & Leamer (1986) and in particular Edgerton & Jochumzen (2003, for example p. 14) report similar results.

Step 1. First derive $E_r[dE(e|X,r)/dr]$. From (1) we have that,

$$\begin{aligned} E(e|X,r) &= 0[\Pr(\epsilon_E > X\beta_E - r\delta_E)] + 1[\Pr(\epsilon_E < X\beta_E - r\delta_E)] \\ &= \Pr(\epsilon_E < X\beta_E - r\delta_E) = F_E(X\beta_E - r\delta_E), \end{aligned} \tag{10}$$

where $F_E(.)$ is the cdf of ϵ_E. It follows directly that,

$$dE(e|X,r)/dr = -\delta_E f_E(X\beta_E - r\delta_E), \tag{11}$$

where $f_E(.)$ is the pdf of ϵ_E. Taking expectation over r, we obtain,

$$E_r[dE(e|X,r)/dr] = -E_r[\delta_E f_E(X\beta_E - r\delta_E)]. \tag{12}$$

Step 2. We now derive $E(e|X,I)$. First note that by the conditional expectation function decomposition property (Theorem 3.1.1 in Angrist & Pischke, 2009) we can express r as the sum of a deterministic and a stochastic part,

$$r = E(r|X,I) + \tau. \tag{13}$$

where τ is stochastic with $E(\tau|X,I) = 0$. Then transform $E(e|X,r)$ by applying (13),

$$\begin{aligned} E(e|X,r) &= F_E(X\beta_E - r\delta_E) \text{ by (10)} \\ &= F_E(X\beta_E - \delta_E(E(r|X,I) + \tau)) \text{ by (13)} \\ &= E(e|X,I,\tau) \end{aligned} \tag{14}$$

By the law of iterated expectations (e.g., Angrist & Pische, 2009, equation 3.1.1) it must be the case that,

$$E(e|X,I) = E_\tau E(e|X,I,\tau). \tag{15}$$

Step 3. In this case, the derivative of $E(e|X,I)$ with respect to I exists and we can complete the proof by applying standard differentiation techniques.

$$\begin{aligned}
dE(e|X,I)/dI &= d[E_\tau E(e|X,I,\tau)]/dI \text{ by (15)} \\
&= E_\tau[dE(e|X,I,\tau)/dI] \text{ because } E_\tau[.] \text{ is a linear operator} \\
&= E_\tau[d[F_E(X\beta_E - \delta_E(E(r|X,I)+\tau))]/dI] \text{ by (14)} \\
&= -E_\tau[f_E(X\beta_E - \delta_E(E(r|X,I)+\tau)] \cdot \delta_E dE(r|X,I)/dI \text{ by the chain rule} \\
&= -E_r[\delta_E f_E(X\beta_E - \delta_E r)] \cdot dE(r|X,I)/dI \text{ by (13)} \\
&= E_r[dE(e|X,r)/dr)] \cdot dE(r|X,I)/dI \text{ by (12)}
\end{aligned} \tag{16}$$

That proves (9) and hence Remark 1 when the estimation model (4) is derived from (1) and (3).

One can note that this result does not hinge on the participation model (2) being linear. If $I = h(X\beta - r\delta_I - \epsilon_I)$, where $h(.)$ is some differentiable function, then $dE(r|X,I)$ changes shape but the proof goes through in exactly the same manner. This point may be helpful in understanding why the discrete case analyzed below yield similar results to the continuous case.

For I discrete, i.e., given by (3), Step 0 - Step 2 go through as before. Step 3 needs to be modified (the chain rule does not apply as in the fourth line of (16)). We now provide such a modification by applying the discrete chain rule theorem of Chalice (2001) to prove a discrete approximation of (9). First define,

$$\begin{aligned}
\Delta e_I &= E(e|X,I=1) - E(e|X,I=0). \\
\Delta r_I &= E(r|X,I=1) - E(r|X,I=0)
\end{aligned} \tag{17}$$

Δr_I and Δe_I are *forward discrete derivatives* in the terminology of Chalice (2001). Remark 1 states that $e_I = e_r r_I$ which by (17) in the discrete model is equivalent to

$$\Delta e_I = \Delta e_r \Delta r_I \tag{18}$$

where Δe_r is the discrete-world analogue of e_r, to be defined below.

We wish to evaluate,

$$\begin{aligned}\Delta e_I &= E_\tau E(e|X, I=1,\tau) - E_\tau E(e|X, I=0,\tau) \text{ by (15)} \\ &= E_\tau[E(e|X, I=1,\tau) - E(e|X, I=0,\tau)] \text{ because } E(.) \text{ is a linear operator} \\ &= E_\tau[F_E(X\beta_E - \delta_E(E(r|X, I=1)+\tau) - F_E(X\beta_E - \delta_E(E(r|X, I=0)+\tau)] \text{ by (14)}\end{aligned} \tag{19}$$

Chalice (2001) provides a "chain rule" for expressions such as the third line of (19) that is helpful to realize that this expression does indeed boil down to a discrete-world approximation to (9). Define $r_1 = E(r|X, I=1)$ and $r_2 = E(r|X, I=0)$, and scale r so that r_1 and r_2 are integers. Furthermore define $N = r_2 - r_1$, and define the sequence

$$\Delta F_E(k) = E_\tau[F_E(X\beta_E - \delta_E(k+1)+\tau) - F_E(X\beta_E - \delta_E k+\tau)],\ k = 1, 2, ... \tag{20}$$

$\Delta F_E(k)$ is the change in entrepreneurship probability when expected risk aversion increases from k to $k+1$, integrated over τ. We are interested in the per-unit change in entrepreneurship probability when expected risk aversion increases from r_1 to r_2 (which corresponds to I changing from 0 to 1). Applying Theorem 6 of Chalice (2001), the so-called chain rule for sequences, we have that,

$$\Delta e_I = \Delta e_r \Delta r_I \tag{21}$$

where $\Delta e_r = \frac{1}{N}\sum_{k=r_1}^{r_2-1}\Delta F_E(k)$. Thus Δe_r is the average of the N forward derivatives starting at r_1 and ending at r_2, integrated over τ. In other words Δe_r is the per-unit change in $E(e|X, r, \tau)$ when expected risk aversion increases from r_1 to r_2 (which corresponds to I changing from 0 to 1). Note that by defining a sufficiently fine discrete grid (i.e., scaling of r) the approximation underlying (21) should come close to being exact.

9.2 Proof of Remark 2

Step 1. Recall from (8) that an individual becomes entrepreneur if,

$$EU(y, r) = \alpha + X\beta_y - r\sigma_\tau^2 > w. \tag{22}$$

Define

$$\bar{\alpha} = \alpha : EU(y,r) - w = 0,$$

Substituting into (22), we obtain,

$$\bar{\alpha} = w - X\beta_y + r\sigma_\tau^2. \tag{23}$$

We see immediately that,

$$\frac{\partial \bar{\alpha}}{\partial r} = \sigma_\tau^2 > 0$$

This means that a more risk averse individual requires a lower fixed cost of entry in order to be willing to enter entrepreneurship.

Step 2. Note that expected entrepreneurial income, conditional on entry, equals,

$$\begin{aligned} E(y|X,r,e &= 1) = \beta_y X + \frac{1}{p}\int_{\bar{\alpha}}^{\infty} \alpha f_{\alpha,X}(\alpha) d\alpha + E(\epsilon_y|X,r,e=1) \qquad (24) \\ &= \beta_y X + \frac{1}{p}\int_{\bar{\alpha}}^{\infty} \alpha f_{\alpha,X}(\alpha) d\alpha \end{aligned}$$

where $p = \int_{\bar{\alpha}}^{\infty} f_{\alpha,X}(\alpha)d\alpha$ and $dp/dr = -\frac{\partial \bar{\alpha}}{\partial r} f_{\alpha,X}(\bar{\alpha}) < 0$.

Step 3. Differentiate (24) with respect to r to obtain,

$$\begin{aligned} dE(y|X,r,e &= 1)/dr = -\frac{1}{p}\frac{\partial \bar{\alpha}}{\partial r}\bar{\alpha} f_{\alpha,X}(\bar{\alpha}) - \frac{dp/dr}{p^2}\int_{\bar{\alpha}}^{\infty} \alpha f_{\alpha,X}(\alpha)d\alpha \qquad (25) \\ &= -\frac{\partial \bar{\alpha}}{\partial r}\frac{f_{\alpha,X}(\bar{\alpha})}{p}[\bar{\alpha} - \frac{1}{p}\int_{\bar{\alpha}}^{\infty} \alpha f_{\alpha,X}(\alpha)d\alpha]. \end{aligned}$$

Because the term in brackets is negative, this expression is positive.

10 Appendix B: Model with constant absolute risk aversion

Here we show that our reduced-form model of entrepreneurial entry, (1), can be derived from an underlying model where individuals have constant absolute risk aversion r and realized entrepreneurial returns are normally distributed. We also show that Remark 2 holds under these assumptions. Suppose first that entrepreneurial income y is given by,

$$y = \alpha_i + X_i \beta_y + \epsilon_{i,y}, \tag{26}$$

where α_i and X_i are known to the agent and $\epsilon_{i,y}$ is unknown. α_i is stochastic with density $f_{\alpha,X}$. We assume that that $\epsilon_{i,y}$ is normally distributed with mean zero and variance σ_y^2. Furthermore suppose that wage income w is given by,

$$w_i = \alpha_{i,w} + X_i \beta_w + \epsilon_{i,w}, \tag{27}$$

where $\epsilon_{i,w}$ is normally distributed with mean zero and variance σ_w^2. We assume that wage work is less risky than entrepreneurship, so that $\sigma_w^2 < \sigma_y^2$. Furthermore, we assume that utility is exponential with $U(y) = -\exp(-ry)$, where r is the coefficient of absolute risk aversion, where a lower r means less risk averse. It is well-known that expected utility, under these assumptions, is separable in expected income and risk and can be written as,

$$\begin{aligned} E(U(y)) &= \alpha_y + X\beta_y - r\sigma_y^2, \text{ and} \\ E(U(w)) &= X\beta_w - r\sigma_w^2 \end{aligned} \tag{28}$$

The final term in each expression is the "risk cost". An individual makes entrepreneurial entry if $e = 1$,

$$e = \mathbf{1}[E(U(y)) - E(U(w)] = \mathbf{1}(\alpha_E + X\beta_E - r\delta_E - \epsilon_E) \tag{29}$$

Substituting in for $\alpha_E = \alpha_y - \alpha_w$, $\beta_E = \beta_y - \beta_w$ and $\delta_E = \sigma_y^2 - \sigma_w^2$, we obtain (1). It is straightforward and hence omitted that Remark 2 holds for the model defined by (26) and (29); the proof follows along the same lines as the proof in the text.

11 Appendix C: Additional output

Here we present additional descriptive statistics and regression results.

11.1 Entry and activity by industry and year

Table C1 here

11.2 Robustness: Alternative methodologies

Table C2 here

11.3 Robustness: Sub-samples, exclusions, and liquidity checks

Table C3 here

11.4 Start-up size

Table C4 here

11.5 Sensation-seeking

Table C5 here

11.6 Correlation matrix

Table C6 here

11.7 Event timeline

Figure C1 here

11.8 Entry versus wealth

Figure C2 here

12 Appendix D. Description of matching

To accommodate that individuals with low wealth or income are unlikely to invest in the stock market or start a firm, we include a large number of sociodemographic controls, including education, previous income, and measures of personal wealth. As another way to deal with selection, we use regression weights, where the regression weights are calculated using two nearest-neighbor propensity score matching. Here we describe the matching procedure and sample selection for the results in Table C2, Panel A and Panel C in more detail.

The outcome variable in the first stage regression is a stock ownership dummy variable. The controls are the socio-demographic characteristics used in Table 2, column 3. The matching estimates are shown in full in Column 1 of Table D1. The idea of propensity score matching is to match stock market investors with individuals who do not invest in the stock market, but whose ex ante probability of investing in the stock market – as predicted by their pre-treatment characteristics – is 'identical' (see Rosenbaum and Rubin, 1983).[33] We impose a caliper (i.e., radius) of 0.05, i.e., stock market investors that have no comparison individual and whose estimated propensity score is within 0.05 of their own estimated propensity score are discarded to avoid bad matches. Imposing this caliper, we only lose 8 of the 68,803 stock market investors in the sample. We select two nearest neighbors and impose the common support criterion (the results are robust to other orders of matching, e.g. one nearest neighbour, drawing controls without replacement). Before matching we have 397,019 individuals in total, of which 68,803 are investors. After matching, we retain 68,795 investors and 93,486 controls.

Table D1 here

Table 1, Panel A shows summary statistics for selected control and outcome variables, before and after matching. While most of the differences in the means of the control variables between investors and non-investors are statistically significant prior to matching (Column 4), after matching the differences in averages become smaller and statistically

[33] We use a version of Leuven and Sianesi's (2003) Stata module psmatch2 (2010, version 4.0.4, http://ideas.repec.org/c/boc/bocode/s432001.html) to perform propensity-score matching.

insignificant for the majority of the control variables (e.g. household wealth and salary income, as shown in Column 7). Another way to look at the question of whether matching eliminates differences is to run a probit regression on the sample consisting of investors and matched non-investors only (using the regression weights obtained from the propensity score matching). The results are reported in Column 2 of Table D1. The table shows that the vast majority of the control variables become statistically insignificant and that the pseudo-R^2 drops from 0.108 in Column 1 to 0.0005 in Column 2. It thus seems that matching to a large extent eliminates ex-ante differences between investors and non-investors.

13 Appendix E: relation between r_I and σ_I^2

The variation in the proxy I in (4) conditional on controls is jointly determined by variation in risk tolerance and in additional (non-risk tolerance related) variation, captured by ϵ_I in (2) and (3). A larger σ_I^2 means that the proxy becomes more weakly related to risk tolerance. In the text, we claim that $|r_I|$ tends to become smaller (and e_I a more conservative estimate of e_r) when σ_I^2 increases. Here we substantiate this claim. Throughout the proof we assume for convenience that $r > 0$, i.e., that the individuals are risk averse (the proof easily generalizes to risk-neutrality and risk-loving preferences).

Continuous case. We consider $E(r|I,X)$ when I is generated by (3). We first analyze the simpler case when X is dropped from (3), i.e., $I = -r\delta_I - \epsilon_I$. The same results apply when r and X are independent. By standard formula (e.g., Angrist & Pischke, 2009, Theorem 3.1.4),

$$dE(r|I)/dI = \frac{cov(r,I)}{var(I)} \tag{30}$$

The right hand side of (30) is just the population analogue of the OLS estimator when regressing r on I, i.e., the "reverse regression" (sometimes also referred to as the "inverse regression" in the literature). Observe that by the independence of r and ϵ_I,

$$\begin{aligned} var(I) &= var(-r\delta_I - \epsilon_I) = var(r\delta_I + \epsilon_I) \\ &= \delta_I^2\sigma_r^2 + \sigma_I^2, \end{aligned} \tag{31}$$

and,

$$
\begin{aligned}
cov(r, I) &= cov(r, -r\delta_I - \epsilon_I) \\
&= -cov(r, r\delta_I) - cov(r, \epsilon_I) \\
&= -\delta_I \sigma_r^2.
\end{aligned}
\tag{32}
$$

Substituting (31) and (32) into (30) we obtain,

$$
r_I = dE(r|I)/dI = \frac{-\delta_I \sigma_r^2}{\delta_I^2 \sigma_r^2 + \sigma_I^2}
\tag{33}
$$

It follows that $d|r_I|/d\sigma_I^2 < 0$ and that $r_I \to 0$ as $\sigma_I^2 \to \infty$. When $\sigma_I^2 = 0$, i.e., a deterministic relationship between r and I *exists*, then $dE(r|I)/dI = -1/\delta_I$ (this result obviously also follows from differentiating (3) after substituting in $\epsilon = 0$. We can note that (33) bears a close resemblance to standard results for measurement error models (e.g., Wooldridge, 2006, equation 4.48).

Let us now consider $E(r|I, X)$. We can use the regression anatomy formula (Angrist & Pischke, 2009, equation 3.1.3) to get,

$$
r_I = dE(r|X, I)/dI = \frac{cov(r, \tilde{I})}{var(\tilde{I})}
\tag{34}
$$

where $\tilde{I}$ is the residual from a regression of I on X. Hence r_I is a bivariate slope coefficient for I after partialing out the effect of X. Using a very similar procedure to Wooldridge (2006, equation 4.47) it can be shown that also in this case r_I is monotonic in σ_I^2 and that $r_I \to 0$ as $\sigma_I^2 \to \infty$.

Discrete case. Let us now consider (2). We start out with the case where X is dropped, and discuss its role at the end. Suppose that $\epsilon = k\psi$, where $k > 0$ is a constant and ψ is a random variable with zero mean and finite variance σ_ψ^2. Increasing k induces distributions of ϵ that are dominated by the second order stochastic dominance criterion, and $d\sigma_I^2/k > 0$. We show that (i) $dE(r|I)/dI < 0$ and that (ii) $dE(r|I)/dI \to 0$ as $k \to \infty$.

Because I is a dummy variable, $E(r|I)$ must be linear in I, and as in the continuous case we have that $dE(r|I)/dI = \dfrac{cov(r,I)}{var(I)}$. Note that $E(I|\epsilon) = \Pr(I = 1|\epsilon) = \Pr(r < -\epsilon/\delta_I)$, and define $P = E(I|\epsilon)$. By the law of total variance,

$$\begin{aligned} var(I) &= E[var(I|\epsilon)] + var[E(I|\epsilon)] \\ &= E[P(1-P)] + var(P) \\ &= E[P(1-P)] + E[(E(P)-P)^2] \\ &= E[P - P^2 + \bar{P}^2 - 2\bar{P}^2 + P^2] \\ &= E[P - \bar{P}^2] = \bar{P}(1-\bar{P}) \end{aligned} \tag{35}$$

where $\bar{P} = E(P) > 0$. Let us now turn to $cov(r, I)$. First note that,

$$\begin{aligned} cov(r,I) &= E[cov(r,I|\epsilon)] + cov[E(r|\epsilon), E(I|\epsilon)] \\ &= E[cov(r,I|\epsilon)] \end{aligned} \tag{36}$$

where expectation is taken over ϵ. The first line in (36) is sometimes refered to the as the law of total covariance. The second line follows from r being independent of ϵ, and therefore $E(r|\epsilon)$ being a constant. Define $\bar{r} = E(r)$ and note that $E(r|\epsilon) = E(r)$ by the independence of r and ϵ. Therefore,

$$\begin{aligned} E[cov(r,I|\epsilon)] &= E[E(rI|\epsilon) - E(r|\epsilon)\cdot E(I|\epsilon)] \text{ by definition} \\ &= E[E(rI|\epsilon) - \bar{r}P] \text{ because } E(r|\epsilon) = E(r) \end{aligned} \tag{37}$$

Using the law of total expectation on $E(rI|\epsilon)$,

$$\begin{aligned} E(rI|\epsilon) &= E(rI|\epsilon, I=1)P + E(rI|\epsilon, I=0)(1-P) \\ &= E(rI|\epsilon, I=1)P \text{ because } rI = 0 \text{ when } I = 0 \\ &= E(r|\epsilon, I=1)P \text{ because } rI = r \text{ for } I = 1 \\ &= E(r|r < -\epsilon/\delta_I)P. \end{aligned} \tag{38}$$

Substituting (37) and (38) into (36), we obtain

$$\begin{aligned} cov(r,I) &= E[cov(r,I|\epsilon)] = E[E(rI|\epsilon) - \bar{r}P] \\ &= E[[E(r|r < -\epsilon/\delta_I) - \bar{r}]P] \end{aligned} \tag{39}$$

Putting together (35) and (39),

$$dE(r|I)/dI = \frac{cov(r,I)}{var(I)} = \frac{E[[E(r|r < -\epsilon/\delta_I) - \bar{r}]P]}{\bar{P}(1-\bar{P})} \tag{40}$$

Clearly $E(r|r < -\epsilon/\delta_I) < E(r) = \bar{r}$. and hence $dE(r|I)/dI < 0$. That proves (i).

To prove (ii), we can rewrite (40) by observing that $E(r|r < -\epsilon/\delta_I) = E[\int_0^{-\epsilon/\delta_I} rf(r)dr]$ where $f(.)$ is the density function of r, to obtain,

$$dE(r|I)/dI = \frac{E[\int_0^{-\epsilon/\delta_I} rf(r)dr - P\bar{r}]}{\bar{P}(1-\bar{P})};\ \epsilon < 0 \tag{41}$$

To prove (ii), recall that $\epsilon = k\psi$ where k is a constant and ψ is a mean zero random variable. We now investigate the limit of $dE(r|I)/dI$ as $k \to \infty$. When $k \to \infty$ then $\bar{P} \to \frac{1}{2}$ and the denominator of (41) converges to $\frac{1}{4}$. Consider the numerator. Let $g_\epsilon(.)$ and $g_\psi(.)$ be the density function of ϵ and ψ, respectively, and let $f(.)$ be the density function of r. For any $\psi, \epsilon < 0$ we must have the following,

$$\begin{aligned} \lim_{k\to\infty} cov(r,I) &= \lim_{k\to\infty} E[\int_0^{-\epsilon/\delta_I} rf(r)dr - P\bar{r}] \\ &= \lim_{k\to\infty} \int_0^\infty [\int_0^{-\epsilon/\delta_I} rf(r)dr - P\bar{r}]g_\epsilon(\epsilon)d\epsilon \\ &= \int_0^\infty \{\lim_{k\to\infty} \int_0^{-k\psi/\delta_I} rf(r)dr - \lim_{k\to\infty}(P\bar{r})\}g_\psi(\psi)d\psi \\ &= \int_0^\infty \{\int_0^\infty rf(r)dr - \bar{r}\}g_\psi(\psi)d\psi \text{ because } \lim_{k\to\infty}(P) = 1 \\ &= \int_0^\infty \{\bar{r} - \bar{r}\}g(\psi)d\psi = 0. \end{aligned} \tag{42}$$

Thus,

$$\lim_{k\to\infty} dE(r|I)/dI = \frac{0}{1/4} = 0 \tag{43}$$

That proves (ii).

In order to investigate the non-linear case in more detail, we have simulated (2) when X and r are correlated normally distributed variables, and ϵ_I is independently normally distributed. We have numerically confirmed, for a large range of parameter values, that r_I monotonically tends toward zero as σ_I^2 increases.

Table 1
Summary Statistics: Averages and Mean Differences

Panel A: Individual characteristics

Sample	Unmatched				Matched		
	Full sample	*Investors*	*Non-Inv.*	*Diff.*	*Investors*	*Non-Inv.*	*Diff.*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Entrepreneur	1.59%	2.88%	1.32%	1.56%***	2.88%	2.05%	0.82%***
Investor	17.33%	100.00%	0.00%	100.00%	100.00%	0.00%	100.00%
Age	37.93	39.18	37.67	1.51***	39.18	39.26	-0.08**
Years of education	12.10	12.88	11.93	0.95***	12.88	12.91	-0.03**
Business education	14.91%	21.66%	13.49%	8.17%***	21.65%	21.74%	-0.09%
Single	27.98%	24.18%	28.77%	-4.59%***	24.18%	23.96%	0.22%
Household wealth	232,140	335,587	210,455	125,133***	335,425	333,093	2,333
Personal Wealth	306,633	446,459	277,321	169,138***	446,288	427,463	18,825***
Net wealth	-94,608	-19,550	-110,342	90,792***	-19,624	-53,790	34,166***
Net capital income	-27,845	-22,491	-28,968	6,477***	-22,504	-27,523	5,018***
Salary income	266,794	316,397	256,396	60,001***	316,347	316,054	293
Debt to wealth ratio	2.126	1.654	2.225	-0.571***	1.654	1.718	-0.064***
Debt to income	1.497	1.443	1.508	-0.065***	1.443	1.504	-0.061***
Self-employed in the past	6.42%	7.81%	6.12%	1.68%***	7.80%	8.18%	-0.38%***
Parent self-employed	10.64%	11.95%	10.38%	1.57%***	11.95%	10.87%	1.08%***
Parent investor	29.65%	92.77%	16.80%	75.97%***	92.77%	19.50%	73.27%***
High-tech entry	0.11%	0.24%	0.08%	0.15%***	0.24%	0.15%	0.09%***
Medium-high-tech entry	0.72%	1.59%	0.54%	1.05%***	1.59%	1.04%	0.55%***
Car horse power	95.91	100.38	94.76	5.62***	100.38	99.47	0.91
Car space (cm^2)	81,730	81,542	81,778	-236	81,541	82,142	-601
No. of observations	*397,019*	*68,803*	*328,216*		*68,795*	*93,486*	

Panel B: Stock market variables

	All investors	*Entry*	*No Entry*	*Diff.*
	(8)	(9)	(10)	(11)
Number of trades	2.856	5.711	2.771	2.940***
Number of stocks	2.002	2.428	1.990	0.438***
One stock holder	49.43%	38.94%	49.74%	-10.80%***
St. Dev. monthly returns	0.0884	0.0926	0.0883	0.0044***
Years active	3.67	3.54	3.68	-0.14***
# observations	*68,803*	*1,980*	*66,823*	

Panel C: Firm variables

	All firms	*Investors*	*Non-Inv.*	*Diff.*
	(12)	(13)	(14)	(15)
Equity	241,580	272,913	227,243	45,670
Sales	3,848.2	3,663.3	3,934.1	-270.8
EBITDA	324.76	344.82	315.41	29.42***
OROA	0.0990	0.0896	0.1034	-0.0138***
St.Dev. $_{OROA}$	0.2668	0.2658	0.2685	
Debt	2,910.6	5,939.0	1,503.6	4,435.4**
Debt-to-assets	83.81%	81.20%	85.03%	-3.83%
Debt-to-equity	10.75%	12.96%	9.73%	3.22%
Num. employees	2.72	2.41	2.87	-0.46***
Survival -2 years	67.34%	61.77%	69.89%	-8.12%***
Survival -3 years	62.75%	58.46%	64.71%	-6.24%***
Survival -4 years	58.82%	54.54%	60.80%	-6.27%***
Survival -8 years	48.96%	44.86%	50.88%	-6.02%***
# observations	*6,307*	*1,980*	*4,327*	

Notes: This table reports averages and mean differences for the main variables. Panel A presents average individual characteristics, for: the full sample in column 1; unmatched investors and non-investors in columns 2 and 3; matched investors and non-investors in columns 5 and 6. Panel B presents averages for the stock market variables in the whole investor sample in column 8; for entrepreneurs and non-entrepreneurs in columns 9 and 10. Panel C presents means the main firm level variables for the start-up year in column 12, with the exception of EBITDA and OROA, for which averages from the panel sample of firm activity are reported (43,917 observations); means for investors and non-investors are presented in columns 13-14. Columns 4, 7, 11 and 15 report differences in averages and the stars denote levels of significance from a t-test of mean differences between the two groups compared (* $p<0.10$, ** $p<0.05$, *** $p<0.01$).

Table 2
Entry Regressions
Linear Probability Models

Panel A:

Dep. Var.:	E	E	E	E	E	SE	HT	MHT	$CCM^{\perp}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Investor	0.0096	0.0082	0.0078	0.0080	0.0068	0.0048	0.0008	0.0050	1.4948
	[14.21]***	[12.08]***	[11.56]***	[9.19]***	[6.23]***	[4.54]***	[4.08]***	[10.20]***	[10.44]***
Log(Household wealth in 1993-94)	0.0025	-0.0117	-0.2243	0.0305	0.0300	0.6485	0.0512	-0.1822	0.000001
	[15.15]***	[9.95]***	[1.09]	[0.12]	[0.12]	[1.46]	[0.92]	[1.36]	[0.81]
[Log(House. wealth in 1993-94)]^2	-	0.0007	0.0487	-0.0100	-0.0098	-0.1482	-0.0096	0.0403	98.3544
		[11.32]***	[1.08]	[0.18]	[0.18]	[1.55]	[0.80]	[1.36]	[1.13]
[Log(House. wealth in 1993-94)]^3	-	-	-0.0052	0.0013	0.0013	0.0159	0.0009	-0.0044	0.5776
			[1.06]	[0.21]	[0.21]	[1.58]	[0.69]	[1.37]	[1.31]
[Log(House. wealth in 1993-94)]^4	-	-	0.0003	-0.0001	-0.0001	-0.0008	-0.00004	0.0002	1.0315
			[1.03]	[0.26]	[0.26]	[1.62]	[0.56]	[1.38]	[1.48]
[Log(House. wealth in 1993-94)]^5	-	-	-0.00001	0.00001	0.00001	0.00001	0.000001	-0.00001	0.9993
			[0.96]	[0.32]	[0.32]	[1.68]*	[0.43]	[1.38]	[1.62]
Parent self-employed	-	-	-	0.0048	0.0048		-	-	-
				[5.01]***	[5.02]***				
Parent investor	-	-	-	-	0.0015		-	-	-
					[1.83]*				
Linear prediction	*0.0159*	*0.0159*	*0.0159*	*0.0135*	*0.0135*	0.0604	*0.0011*	*0.0072*	*0.0476*
% Investor effect	60.34%	51.40%	49.35%	59.31%	50.36%	7.99%	74.04%	70.00%	-
No. of Observations	397,019	397,019	397,019	237,933	237,933	397019	397,019	397,019	91,224
Adjusted R^2	0.019	0.021	0.021	0.016	0.016	0.141	0.003	0.016	-

Panel B: Alternative risk proxies:

Dep. Var.: E	Log (Portf. val. to wealth)		Log (Portf. val. to income)		Log (Debt to wealth)	Log (Debt to income)
	(10)	(11)	(12)	(13)	(14)	(15)
Risk proxy	0.0022	0.0018	0.0019	0.0018	0.0013	0.0014
	[6.66]***	[5.04]***	[5.99]***	[5.15]***	[14.56]***	[15.72]***
Non-Investor	-0.0013	-	-0.0020	-	-	-
	[1.19]		[1.88]*			
Linear prediction	*0.0159*	*0.0288*	*0.0159*	*0.0288*	*0.0159*	*0.0159*
% proxy effect	53.94%	28.65%	51.64%	31.59%	11.46%	10.92%
No. of Observations	397,019	68,803	397,019	68,803	397,019	397,019
Adjusted R^2	0.021	0.024	0.021	0.024	0.021	0.021

Panel C: Alternative wealth proxies:

Dep. Var.: E	Log(Personal Wealth)	Net Capital Income	Net Wealth
	(16)	(17)	(18)
Investor	0.0077	0.0090	0.0094
	[11.35]***	[13.29]***	[13.84]***
Linear prediction	*0.0159*	*0.0159*	*0.0159*
% Investor effect	48.50%	56.81%	59.32%
No. of Observations	397,019	397,019	397,019
Adjusted R^2	0.021	0.022	0.021

Notes: The table reports estimates of entry regressions for new firms outside the financial and real estate sector, incorporated between 2000 and 2007, and where an individual owns at least 50 percent at the start-up date (denoted by *E*). In Panel A, columns 1-8 report coefficients and t-statistics in brackets from linear probability models with robust standard errors. The sample is composed of 1995-1999 investors and non-investors. Pre-determined values of the control variables are used as averages across 1993/94. In column 1, log household wealth is entered linearly; in column 2 it is entered as a 2nd order polynomial, and as a 5th order polynomial in column 3. Column 4 also incorporates a 2nd order polynomial in log parental wealth, along with parental and spousal self-employment and business education dummies. The estimates are for the sub-sample of individuals whose parents are in the labor force during 1993 - 1999 (*i.e.* observations are dropped for individuals whose parents were deceased or retired). Column 5 adds parental investor status to the specification in column 4. Column 6 replicates the estimates of column 3, having new self-employment entry between 1995 and 2007 as the dependent variable (denoted by *SE*). Columns 7 and 8 present models that have entrepreneurial entry in high-tech (denoted by *HT*) and medium-high-tech industries (denoted by *MHT*), respectively, as dependent variable. Column 9 reports odds ratios and z-statistics for the likelihood of entry into entrepreneurship, utilizing a conditional logit model (denoted by *CCM*). The sample is composed of 4,344 new entrepreneurs starting up between 2000 and 2007 (cases), each of which is matched with 20 non-entrepreneurs (controls). 1-20 case-control matching is used based on 3 age groups, 3 education groups, 3 marital status groups, household wealth centile and salary income centile (675 categories). The following pre-determined variables are used as controls in all regressions: a 2nd order polynomial in age, years of education, marital status, the logarithms of the number of children and the number of siblings, 3rd order polynomial in salary income and an interaction term between the logs of household wealth and salary income (with the exception of columns 1 and 2, in which linear and quadratic log salary is used, respectively), region of residence, type of education (10 groups), 2-digit industry of employment (SIC) codes, and five firm size (number of employees) dummy variables. Panel B presents entrepreneurial entry models in which the risk proxy is the fraction of wealth invested in the stock market (columns 10-13) and personal leverage in columns 14 and 15. The specification is the same as column (3), replacing the investor variable with log(portfolio value to wealth ratio) in columns 10 and 11. Column 10 presents estimates using the full sample of investors and non-investors, replacing missing log(portfolio value to wealth ratio) with zero and introducing a dummy variable for non-investor status. Column 11 presents estimates using the sub-sample of investors. Columns 12 and 13 replicate the estimates of columns 11 and 12, respectively, using log(portfolio value to salary income ratio) as the risk proxy. Columns 14 and 15 replicate the estimates of column 11 using the full sample of investors and non-investors, and having log(debt to wealth ratio) and log(debt to income ratio) as the risk proxies, respectively. Panel C replicates the entrepreneurial entry estimates of column 3, replacing log household wealth with: log individual wealth in column 16; net capital income divided by 1 million in column 17, and net wealth divided by 10 million in column 18. Star levels next to the brackets indicate the levels of significance of the estimates, denoting: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. The "% Investor effect" in Panels A and C is calculated as the ratio of the coefficient divided by the linear prediction (and multiplied by 100). The "% proxy effect" in Panel B is calculated as the percentage increase in the linear prediction induced by an interquartile increase at the level of the risk proxy.

Table 3
Firm Performance

Dep. var.:	OROA	Log (Sales)	Log (Employees)	4-year Survival
Panel A: Investor				
	(1)	(2)	(3)	(4)
Investor	-0.0120	-0.2501	-0.3155	-0.0255
	[2.36]**	[4.12]***	[3.34]***	[1.83]*
Linear prediction	*0.0942*	*6.3624*	*-1.8147*	*0.5978*
% Investor effect	-12.71%	-22.27%	-27.38%	-4.27%
No. of Observations	38,507	38,507	38,507	5,619
Adjusted R^2	0.043	0.172	0.152	0.074
Panel B: Leverage				
	(5)	(6)	(7)	(8)
Log(Debt-to-Income)	-0.0084	-0.0313	-0.0619	-0.0190
	[5.95]***	[3.73]***	[2.33]**	[4.38]***
Linear prediction	*0.0942*	*6.3624*	*-1.8147*	*0.5978*
% proxy effect	-8.36%	-2.96%	-5.76%	-2.95%
No. of Observations	38,507	38,507	38,507	5,619
Adjusted R^2	0.045	0.170	0.151	0.076

Notes: The table reports estimates of performance regressions for new firms outside the financial and real estate sector, incorporated between 2000 and 2007, and where an individual owns at least 50 percent at the start-up date. The entrepreneur characteristics are those reported in the previous two tables. The sample comprises of firms for the years 2000-2010. The panel sample of observations is used in columns 1-3 and 5-7, while a sample with one observation per firm is used in columns 4 and 8 capturing survival in business for at least four years. Columns 1-8 of the two panels report coefficients and t-statistics in brackets from linear regression models with robust standard errors. When the panel sample is used, standard errors are clustered at the firm level to account for serial correlation between repeated observations by the same firm. The following pre-determined variables (entrepreneur characteristics in 1993/94) are used as controls in all regressions: a 2nd order polynomial in age, years of education, marital status, the logarithms of the number of children and the number of siblings, 5th order polynomial in household wealth, 3rd order polynomial in salary income, an interaction term between the logs of household wealth and salary income, region of residence, type of education (10 groups), and five firm size (number of employees) dummy variables. The specifications also incorporate 2-digit start-up industry (NACE) codes, the logarithm of start-up equity and its square. When the panel sample of observations is used, the specifications also include year fixed effects and firm age dummies. In the survival regressions, year of entry dummies are included. Panel B reports estimates in which the logarithm of the debt to salary income ratio is the risk proxy. The latter estimates are robust to the use of the logarithm of debt to wealth ratio as a proxy (available upon request). Star levels next to the brackets are levels of significance of the estimates, denoting: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. The "% Investor effect" is calculated as the ratio of the coefficient divided by the linear prediction for OROA and survival. The calculation of the effect of dummy variables (Panel A) in models with log-transformed dependent variables is based on the formula: $100(\exp(\text{Coef.}-(\text{S.E.}^2/2))-1)$. The "% proxy effect" in Panel B is calculated as the percentage increase in the linear prediction induced by an interquartile increase at the level of the risk proxy.

Table 4
Optimism and Overconfidence

Dep. var.:	Entry	OROA	Log (Sales)	Log (Employees)	4-year Survival
Panel A: Trades					
	(1)	(2)	(3)	(4)	(5)
Log(# of Trades)	0.0087	-0.0005	-0.0379	-0.0284	0.0010
	[6.62]***	[0.32]	[1.97]**	[0.94]	[0.22]
Linear prediction	*0.0288*	*0.0830*	*5.8673*	*-2.4459*	*0.5516*
No. of Observations	68,803	12,415	12,415	12,415	1,781
Adjusted R^2	0.026	0.040	0.175	0.146	0.077
Panel B: One stock					
	(6)	(7)	(8)	(9)	(10)
Owner of one stock	-0.0057	0.0052	0.2155	0.1839	0.0371
	[4.15]***	[0.60]	[2.03]**	[1.12]	[1.52]
Linear prediction	*0.0288*	*0.0830*	*5.8673*	*-2.4459*	*0.5516*
No. of Observations	68,803	12,415	12,415	12,415	1,781
Adjusted R^2	0.025	0.040	0.175	0.146	0.078
Panel C: Debt to Income					
	(11)	(12)	(13)	(14)	(15)
Log(Debt-to-income)	0.0029	-0.0078	-0.0685	-0.1082	-0.0209
	[10.18]***	[3.26]***	[2.18]**	[2.35]**	[2.76]***
Linear prediction	*0.0288*	*0.0830*	*5.8674*	*-2.4459*	*0.5515*
No. of Observations	68,803	12,415	12,415	12,415	1,781
Adjusted R^2	0.026	0.041	0.175	0.147	0.080
Panel D: All Proxies					
	(16)	(17)	(18)	(19)	(20)
Log(# of Trades)	0.0077	-0.0001	-0.0259	-0.0163	0.0041
	[5.63]***	[0.02]	[1.25]	[0.51]	[0.86]
Owner of one stock	-0.0028	0.0058	0.1764	0.1644	0.0462
	[1.99]**	[0.62]	[1.55]	[0.95]	[1.80]*
Log(Debt-to-income)	0.0027	-0.0079	-0.0686	-0.1087	-0.0215
	[9.73]***	[3.27]***	[2.17]**	[2.36]**	[2.84]***
Linear prediction	*0.0288*	*0.0830*	*5.8674*	*-2.4459*	*0.5515*
No. of Observations	68,803	12,415	12,415	12,415	1,781
Adjusted R^2	0.027	0.041	0.177	0.148	0.081

Notes: This table reports estimates of entry and performance regressions for new firms as described in the previous tables. The sample comprises of all investors in the years 1995-1999 in the entry models, while it is restricted to the sample of investors who become entrepreneurs in the performance models. One observation per firm is used for the entry and survival regressions, while the panel 2000-2010 sample of observations is used for the estimation of OROA, log(sales), and log(employment). All columns report coefficients and t-statistics in brackets from linear regression models with robust standard errors. When the panel sample is used, standard errors are clustered at the individual level to account for serial correlation between repeated firm observations. The following pre-determined variables (entrepreneur characteristics in 1993/94) are used as controls: a 2nd order polynomial in age, years of education, marital status, the logarithms of the number of children and the number of siblings, 5th order polynomial in household wealth, 3rd order polynomial in salary income, an interaction term between the logs of household wealth and salary income, region of residence, type of education (10 groups), and five firm size (number of employees) dummy variables. The specifications also incorporate the logarithm of portfolio value and its square. With the exception of the entry (E) models, the specifications also incorporate 2-digit start-up industry (NACE) codes, the logarithm of start-up equity and its square. Finally year fixed effects and firm age dummies are included in all models apart from entry and survival. In the survival regressions, year of entry dummies are included. Panel A reports estimates in which the logarithm of the number of trades is use as a proxy for other behavioural effects. Panel B reports estimates in which ownership of one stock only is the main behavioral proxy. Panel C reports estimates in which the logarithm of debt to salary income ratio is the risk proxy for the investor sample. Finally, Panel D reports estimates in which all three proxies are incorporated simultaneously. Star levels next to the brackets are levels of significance of the estimates, as in the previous tables.

Table C1
Entry and survival by industry and year

Industry (1-digit NACE code)			Entry by year											Activity	
	#Entrants	*(%)*	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Survival	Death
Unknown	220	*(3.49)*	48	44	42	0	59	4	20	3	-	-	-	15	205
Agriculture, hunting, & fishing	90	*(1.43)*	11	16	11	13	11	4	15	9	-	-	-	32	58
Mining and quarrying	26	*(0.41)*	4	4	3	1	2	2	5	5	-	-	-	10	16
Manufacturing	415	*(6.58)*	49	54	56	64	50	34	62	46	-	-	-	207	208
Electricity, gas and water supply	17	*(0.27)*	0	3	1	4	3	0	1	5	-	-	-	3	14
Construction	755	*(11.97)*	67	79	99	84	90	69	143	124	-	-	-	470	285
Wholesale & retail trade, and repairs	1,617	*(25.64)*	217	249	222	235	201	154	204	135	-	-	-	758	859
Hotels and restaurants	183	*(2.90)*	20	21	35	30	25	21	17	14	-	-	-	77	106
Transport, storage and communication	313	*(4.96)*	43	38	30	39	25	37	52	49	-	-	-	144	169
Real estate, renting & business activities	2,283	*(36.20)*	340	270	270	193	234	223	471	282	-	-	-	951	1,332
Education	55	*(0.87)*	6	12	7	2	5	7	10	6	-	-	-	25	30
Health and social work	186	*(2.95)*	22	15	16	14	14	23	51	31	-	-	-	113	73
Other comm., social & personal service	147	*(2.33)*	19	13	20	20	13	18	21	23	-	-	-	62	85
Total # Entrants	6,307	*(100.00)*	846	818	812	699	732	596	1,072	732	-	-	-	2,867	3,440
(% Entrants)			*(13.41)*	*(12.97)*	*(12.87)*	*(11.08)*	*(11.61)*	*(9.45)*	*(17.00)*	*(11.61)*				*(45.46)*	*(54.54)*
Total # Deaths	3,440	*(54.54)*	-	96	195	224	251	332	423	638	535	338	408	-	-
(% Deaths)				*(2.79)*	*(5.67)*	*(6.51)*	*(7.30)*	*(9.65)*	*(12.30)*	*(18.55)*	*(15.55)*	*(9.83)*	*(11.86)*		

Table C2
Robustness: Alternative methodologies

	Risk Proxies							Wealth proxies		
	Investor	*Log(Portfolio value/Wealth)*		*Log(Portfolio value/Income)*		*Log(Debt to wealth)*	*Log(Debt to income)*	*Log(Personal Wealth)*	*Net Cap. Income*	*Net Wealth*
Panel A: LPM Model (Propensity score matching, 1-to-2 nearest neighbor matching, common support)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Risk proxy	0.0081	0.0020	0.0018	0.0019	0.0018	0.0024	0.0024	0.0077	0.0083	0.0083
	[9.18]***	[6.03]***	[5.02]***	[5.68]***	[5.13]***	[11.55]***	[11.98]***	[8.74]***	[9.35]***	[9.41]***
Linear pred.	0.0248	0.0248	0.0288	0.0248	0.0288	0.0248	0.0248	0.0248	0.0248	0.0248
% risk effect	32.60%	35.14%	28.54%	35.25%	31.42%	13.46%	12.17%	31.18%	33.45%	33.47%
No. of obs.	162,152	162,152	68,793	162,152	68,793	162,152	162,152	162,152	162,152	162,152
Adjusted R^2	0.024	0.024	0.026	0.024	0.026	0.024	0.024	0.025	0.025	0.025
Panel B: Probit Model (No matching)										
	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Risk proxy	0.0061	0.0011	0.0018	0.0009	0.0018	0.0020	0.0021	0.0061	0.0074	0.0075
	[6.78]***	[5.66]***	[1.86]*	[4.42]***	[2.87]***	[27.49]***	[19.98]***	[12.67]***	[15.33]***	[15.54]***
Pred. prob.	0.0159	0.0159	0.0288	0.0159	0.0288	0.0159	0.0159	0.0159	0.0159	0.0159
% risk effect	38.57%	31.06%	28.23%	25.76%	30.46%	19.00%	17.06%	38.53%	46.25%	46.98%
No. of obs.	397,019	397,019	68,803	397,019	68,803	397,019	397,019	397,019	397,019	397,019
Pseudo R^2	0.103	0.104	0.088	0.104	0.088	0.104	0.105	0.103	0.105	0.103
Panel C: Probit Model (Propensity score matching, 1-to-2 nearest neighbor matching, common support)										
	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
Risk proxy	0.0083	0.0016	0.0018	0.0015	0.0018	0.0036	0.0036	0.0080	0.0086	0.0085
	[4.74]***	[2.70]***	[3.65]***	[2.86]***	[2.94]***	[3.62]***	[5.71]***	[9.16]***	[9.85]***	[9.71]***
Pred. prob.	0.0248	0.0248	0.0288	0.0248	0.0288	0.0248	0.0248	0.0248	0.0248	0.0248
% risk effect	33.55%	30.17%	28.27%	28.26%	30.51%	22.23%	19.52%	32.34%	34.84%	34.34%
No. of obs.	162,152	162,152	68,793	162,152	68,793	162,152	162,152	162,152	162,152	162,152
Pseudo R^2	0.0922	0.093	0.088	0.093	0.088	0.094	0.094	0.093	0.095	0.092

Notes: Each cell reports the estimated risk proxy coefficient (or average marginal effect) from separate regressions. In panels A-C we replicate columns (3), and (10)-(18) of Table 2 but use other regression methodologies. The regression methodologies are specified in the panel text. For example, in Panel A, we use a weighted linear probability model. Coefficients and t-statistics are reported in brackets (robust standard errors are used). The sample is composed of 1995-1999 investors and matched non-investors. Regression weights from 2 nearest-neighbor propensity score matching are used. In the matching procedure, the outcome variable is a stock ownership dummy and the controls are the socio-demographic characteristics used in column 3 and reported in Column 1 of the Appendix Table D1. In Panel B we use a probit model without matching, and in Panel C we use regression weights as in Panel A, in a probit model for the sample of investors and matched non-investors. In Panels B and C, we report average marginal effects and z-statistics in brackets (robust standard errors are used). The "% risk effect" is calculated as the ratio of the investor coefficient divided by the linear prediction (predicted probability in the non-linear models). The effects reported in parentheses for the other risk proxies capture the effect of an interquartile increase in the risk proxy from the 25th to the 75th percentile of its value. This amounts to an increase from 0.002 to 0.104 in portfolio value to wealth ratio; from 0.017 to 0.113 in portfolio to income ratio; from 0.66 to 2.66 in debt to wealth ratio; and from 0.63 to 2.15 in debt to income ratio.

Table C3
Robustness: Sub-samples, exclusions and liquidity effects

	Coef.	[z]	*Linear pred.*	*%Risk effect*	No. of obs.	Adj. R^2
Panel A: Sub-samples						
(1) Single	0.0114	[8.17]***	*0.0149*	*76.82%*	111,079	0.018
(2) Married	0.0065	[7.87]***	*0.0165*	*39.31%*	249,635	0.022
(3) Widowed/Divorced/Separated	0.0071	[3.19]***	*0.0145*	*49.32%*	36,305	0.024
(4) Age 25-33	0.0152	[9.66]***	*0.0206*	*73.98%*	125,520	0.020
(5) Age 34-42	0.0061	[5.25]***	*0.0169*	*35.94%*	144,528	0.023
(6) Age 43-50	0.0037	[4.21]***	*0.0101*	*36.60%*	126,971	0.020
(7) Wealth 1993-94: 1st quartile (lowest)	0.0107	[6.45]***	*0.0110*	*97.89%*	99,255	0.012
(8) Wealth 1993-94: 2nd quartile	0.0058	[4.55]***	*0.0119*	*49.06%*	99,256	0.014
(9) Wealth 1993-94: 3rd quartile	0.0074	[5.91]***	*0.0152*	*48.59%*	99,254	0.020
(10) Wealth 1993-94: 4th quartile (highest)	0.0081	[6.51]***	*0.0254*	*31.77%*	99,254	0.025
(11) Wealth 1999: 1st quartile (lowest)	0.0084	[5.53]***	*0.0078*	*107.29%*	98,860	0.008
(12) Wealth 1999: 2nd quartile	0.0045	[3.74]***	*0.0097*	*46.11%*	98,878	0.008
(13) Wealth 1999: 3rd quartile	0.0037	[3.36]***	*0.0137*	*27.11%*	98,878	0.010
(14) Wealth 1999: 4th quartile (highest)	0.0046	[3.45]***	*0.0324*	*14.07%*	98,878	0.024
Panel B: Exclusions						
(15) - Business Education	0.0067	[9.42]***	*0.0142*	*47.47%*	337,837	0.019
(16) - Working in small firms (1-10 employees)	0.0077	[11.00]***	*0.0140*	*54.92%*	337,808	0.021
(17) - Oslo/Akershus	0.0068	[9.04]***	*0.0145*	*46.73%*	303,511	0.019
(18) - Unemployed in 1999	0.0077	[11.33]***	*0.0159*	*48.60%*	389,342	0.021
(19) - Father in self-employment	0.0077	[11.28]***	*0.0150*	*51.51%*	373,358	0.021
(20) - Family member in self-employment	0.0074	[10.63]***	*0.0141*	*52.26%*	346,490	0.020
(21) - Self-employment in the past	0.0076	[11.58]***	*0.0133*	*56.95%*	371,545	0.015
(22) - Family business entry	0.0075	[11.47]***	*0.0146*	*51.43%*	396,498	0.020
(23) - Multiple business entry	0.0062	[10.00]***	*0.0134*	*46.37%*	396,000	0.015
(24) - Remaining real estate firms	0.0035	[6.75]***	*0.0102*	*34.59%*	394,736	0.013
(25) - Self-employed	0.0064	[10.61]***	*0.0110*	*58.29%*	367,738	0.014
(26) - Investors with zero trades	0.0103	[12.38]***	*0.0158*	*65.25%*	379,158	0.021
(27) - Investors with zero stock purchases	0.0107	[12.37]***	*0.0158*	*67.71%*	376,419	0.021
Panel C: Liquidity						
(27) Excluding investors with returns>5.48%	0.0128	[9.22]***	*0.0144*	*88.93%*	346,944	0.019
(28) Excluding investors with top quartile returns	0.0080	[10.51]***	*0.0155*	*51.96%*	382,110	0.021
(29) Control for home ownership	0.0078	[11.53]***	*0.0159*	*49.23%*	397,019	0.021
(30) 9th order polynomial in personal wealth	0.0076	[11.21]***	*0.0159*	*47.90%*	397,019	0.021
(31) 7th order polynomial in household wealth	0.0077	[11.41]***	*0.0159*	*48.67%*	397,019	0.021
(32) 5th order polynomial in 1999 personal wealth	0.0047	[6.95]***	*0.0160*	*29.67%*	394,182	0.025
(33) 5th order polynomial in 1999 household wealth	0.0050	[7.39]***	*0.0159*	*31.54%*	395,494	0.024
(34) 5th order pol. in difference in log household wealth btw 1999 and 1993-94	0.0064	[9.49]***	*0.0159*	*40.41%*	395,494	0.022

Notes: Each line reports the estimated coefficient of the investor dummy variable from a separate linear probability regression. The control variables are as in Column 3 of Table 2. All models use robust standard errors, and t-statistics are presented in brackets. In Panel A, each regression is estimated on a different subsample, i.e. the group of individuals specified in the text. In Panel B, the groups of individuals with the characteristic that denoted in the text are dropped from the estimation. This is also the case in lines 26-27 of Panel C. In lines 28-33 of Panel C, we are using the full sample in the estimation of the entry model, adding the variable(s) denoted in the text to the list of control variables. The wealth polynomials in 29-33 replace the polynomial in log(household wealth) in the specification of Column 3, Table 2.

Table C4
Start-up size and Capital structure

Panel A: Start-up size							
Dep. Var.:	**Log(Start-up equity)**						
Other risk proxy (logs)	-	Prtf. value to wealth	Prtf. value to income	Debt to wealth		Debt to income	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Investor	-0.0200	-	-	-	-0.0212	-	-0.0204
	[0.59]				[0.63]		[0.61]
Other risk proxy	-	0.0102	0.0170	-0.0123	-0.0126	-0.0054	-0.0056
		[0.75]	[1.26]	[1.27]	[1.29]	[0.57]	[0.59]
No. of Observations	6,307	1,980	1,980	6,307	6,307	6,307	6,307
Adjusted R^2	0.069	0.095	0.096	0.069	0.069	0.069	0.069
Panel B: Start-up capital structure							
Dep. Var.:	**Log(Debt-to-assets)**						
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Investor	0.0180	-	-	-	0.0208	-	0.0202
	[0.59]				[0.68]		[0.66]
Other risk proxy	-	-0.0177	-0.0206	0.0265	0.0267	0.0237	0.0239
		[1.25]	[1.47]	[2.99]***	[2.99]***	[2.87]***	[2.88]***
No. of Observations	6,004	1,904	1,904	6,004	6,004	6,004	6,004
Adjusted R^2	0.027	0.034	0.034	0.029	0.028	0.028	0.028
Panel C: St.Dev.(OROA) with industry fixed effects							
Dep. Var.:	**St.Dev.(OROA)**						
	(15)	(16)	(17)	(18)	(19)	(20)	(21)
Investor	0.0085	-	-	-	0.0089	-	0.0088
	[2.35]**				[2.45]**		[2.44]**
Other risk proxy	-	-0.0008	-0.0007	0.0034	0.0035	0.0031	0.0032
		[0.54]	[0.48]	[3.42]***	[3.51]***	[3.26]***	[3.35]***
Linear prediction	*0.1830*	*0.1851*	*0.1851*	*0.1830*	*0.1830*	*0.1830*	*0.1830*
No. of Observations	6,050	1,920	1,920	6,050	6,050	6,050	6,050
Adjusted R^2	0.196	0.203	0.203	0.196	0.197	0.196	0.197
Panel D: St.Dev.(OROA) without industry fixed effects							
Dep. Var.:	**St.Dev.(OROA)**						
	(22)	(23)	(24)	(25)	(26)	(27)	(28)
Investor	0.0105	-	-	-	0.0109	-	0.0108
	[2.89]***				[3.00]***		[2.98]***
Other risk proxy	-	-0.0008	-0.0008	0.0037	0.0038	0.0032	0.0033
		[0.55]	[0.56]	[3.63]***	[3.74]***	[3.30]***	[3.41]***
Linear prediction	*0.1830*	*0.1851*	*0.1851*	*0.1830*	*0.1830*	*0.1830*	*0.1830*
No. of Observations	6,050	1,920	1,920	6,050	6,050	6,050	6,050
Adjusted R^2	0.173	0.178	0.178	0.174	0.175	0.173	0.174

Notes: Linear regression models; coefficients and t-statistics in brackets (robust standard errors throughout). Start-up year dummies included throughout. Panel B also incorporates controls for log(start-up equity) and its square. Panels C and D also incorporate controls for log(start-up equity) and its square and dummy variables for maximum firm age. Additional control variables as in Column 3 of Table 2. Panels A-C incorporate controls for start-up industry (2-digit NACE codes), while Panel D excludes these from the specification.

Table C5
Sensation-Seeking

Dep. var.:	Entry	OROA	Log (Sales)	Log (Employees)	4-year Survival
Panel A:					
	(1)	(2)	(3)	(4)	(5)
Log(Car horse power-to-size)	0.0116	-0.0378	-0.2176	-0.6551	-0.0936
	[3.84]***	[1.88]*	[0.84]	[1.80]*	[1.48]
Linear Prediction	*0.0232*	*0.0871*	*6.3420*	*-1.7700*	*0.5953*
% Car power effect	*16.13%*	*-13.26%*	*-1.12%*	*13.02%*	*-5.06%*
No. of Observations	26,665	3,870	3,870	3,870	559
Adjusted R^2	0.023	0.051	0.181	0.177	0.111
Panel B:					
	(6)	(7)	(8)	(9)	(10)
Investor	0.0087	-0.0419	-0.5428	-0.3820	-0.1137
	[3.08]***	[2.58]**	[2.76]***	[1.26]	[2.42]**
Linear Prediction	*0.0232*	*0.0871*	*6.3420*	*-1.7700*	*0.5953*
% Investor effect	*37.69%*	*-48.16%*	*-43.00%*	*-34.80%*	*-19.11%*
No. of Observations	26,665	3,870	3,870	3,870	559
Adjusted R^2	0.023	0.054	0.188	0.176	0.118
Panel C:					
	(11)	(12)	(13)	(14)	(15)
Log(Car horse power-to-size)	0.0114	-0.0354	-0.1865	-0.634	-0.0884
	[3.79]***	[1.81]*	[0.72]	[1.75]*	[1.40]
Investor	0.0086	-0.0409	-0.5372	-0.363	-0.1116
	[3.03]***	[2.51]**	[2.74]***	[1.21]	[2.37]**
Predicted Probability	*0.0232*	*0.0871*	*6.3419*	*-1.7700*	*0.5953*
% Investor effect	*37.05%*	*-46.94%*	*-42.70%*	*-32.50%*	*-18.74%*
No. of Observations	26,665	3,870	3,870	3,870	559
Adjusted R^2	0.023	0.055	0.189	0.179	0.12

Notes: This table reports estimates of entry and performance regressions for new firms as described in the previous tables. The sample comprises of car owners in the year 1999, matched with the full sample of Table 1, Panel A. One observation per firm is used for the entry and survival regressions, while the panel 2000-2010 sample of observations is used for the estimation of OROA, log(sales), and log(employment). All columns report coefficients and t-statistics in brackets from linear regression models with robust standard errors. When the panel sample is used, standard errors are clustered at the firm level to account for serial correlation between repeated firm observations. The following pre-determined variables are used as controls in all regressions: a 2nd order polynomial in age, years of education, marital status, the logarithms of the number of children and the number of siblings, 5th order polynomial in household wealth, 3rd order polynomial in salary income, an interaction term between the logs of household wealth and salary income, region of residence, type of education (10 groups), and five firm size (number of employees) dummy variables. 2-digit industry of employment (SIC) codes are included in the specifications of the entry regressions. Performance regressions incorporate 2-digit start-up industry (NACE) codes, along with the logarithm of start-up equity and its square. Finally year fixed effects and firm age dummies are included in all models apart from entry and survival. In the survival regressions, year of entry dummies are included. Panel A reports estimates in which the logarithm of car horse power divided by size (car length multiplied by width) is used as a proxy for sensation-seeking. Panel B reports estimates in which stock market participation is controlled for in the car owner sub-sample. Panel C reports estimates that incorporate both engine size and investor status. All results are robust when using the car-owner sample of investors and matched non-investors.

Table C6
Pairwise correlation matrix

Panel A: All individuals (*N=397,019*)	Investor	Debt to income	Debt to wealth
Investor	1.0000		
Debt to income	-0.0221 (0.000)	1.0000	
Debt to wealth	-0.0964 (0.000)	0.3309 (0.000)	1.0000

Panel B: Investors (*N=68,803*)	Debt to income	Debt to wealth	Portfolio value to wealth	Portfolio value to income	St. Dev. monthly returns
Debt to income	1.0000				
Debt to wealth	0.4179 (0.000)	1.0000			
Portfolio value to wealth	-0.1033 (0.000)	0.1288 (0.000)	1.0000		
Portfolio value to income	-0.0560 (0.000)	-0.1447 (0.000)	0.7480 (0.000)	1.0000	
St. Dev. monthly returns	0.0240 (0.000)	0.0494 (0.000)	0.0633 (0.000)	0.0014 (0.719)	1.0000

Notes: Levels of significance are given in parentheses

Figure C1
Event timeline

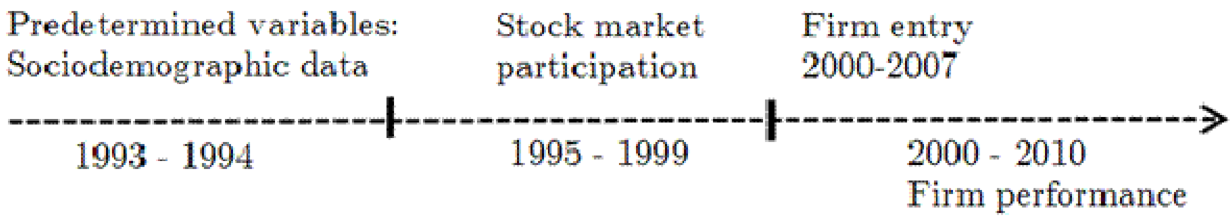
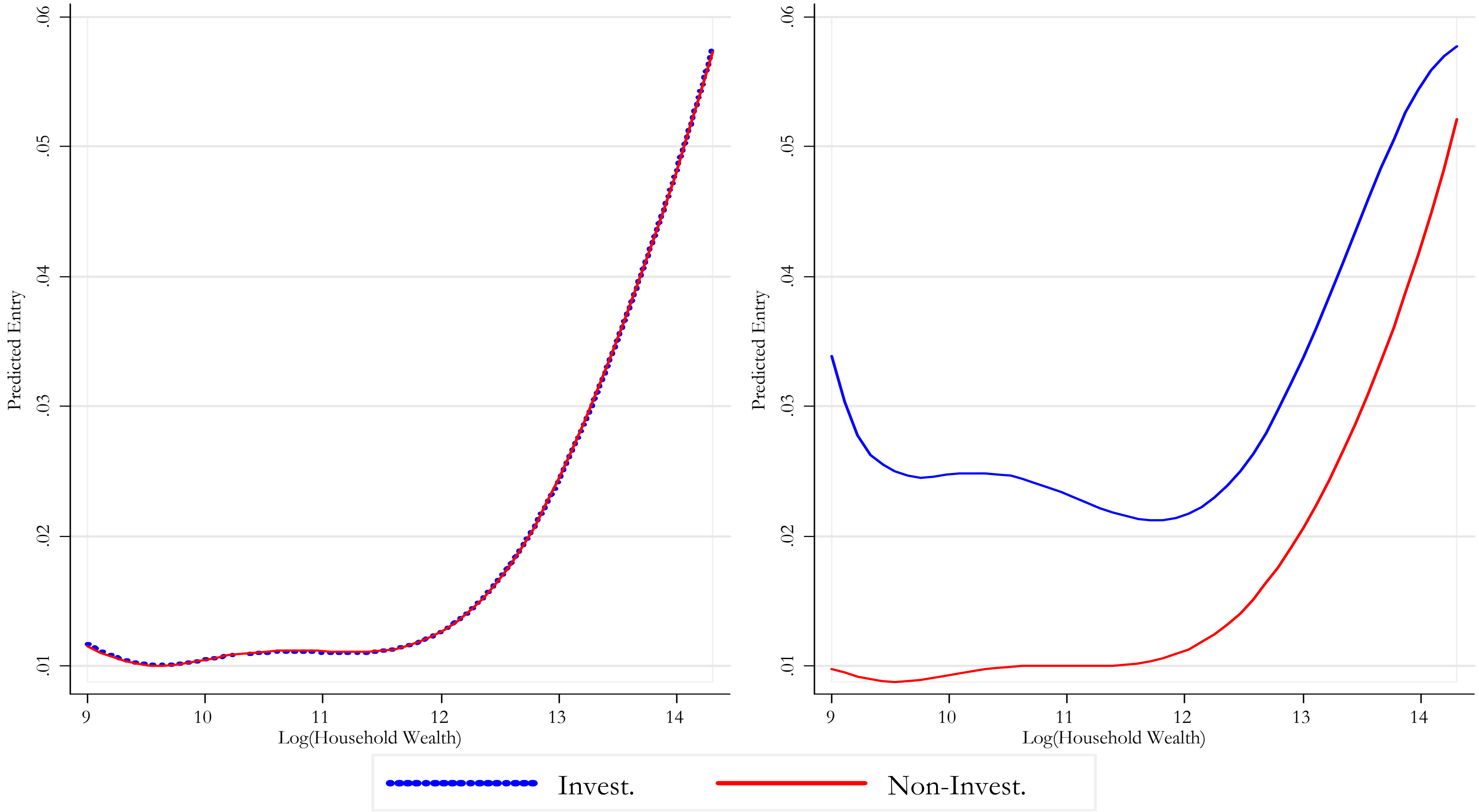


Figure C2
Entry versus wealth



Notes: The figures present 5th order local polynomial regressions of predicted entry on the logarithm of household wealth. The left panel plots predicted values from two separate entry regressions, without and with the investor dummy respectively. The right panel plots predicted values from one single regression for the investor and the non-investor sample respectively (2 separate local polynomial regressions for the predicted values). Predicted values were obtained using linear probability models. The specification in the regressions is that of column 3 in Table 2.

Table D1
Propensity score matching

Dependent variable: Stock ownership 1995-1999 (1/0)				
	(1)		(2)	
Age	-0.0711	[16.68]***	-0.0072	[1.20]
Age squared /1,000	0.9672	[17.37]***	0.0861	[1.11]
Married	0.0344	[4.17]***	-0.0086	[0.74]
Widowed/Separated	-0.0382	[3.42]***	-0.0223	[1.42]
Log(# children)	-0.082	[12.46]***	0.0113	[1.22]
Log(# siblings)	-0.0453	[8.34]***	0.0099	[1.30]
Years of education	0.0488	[36.25]***	-0.0039	[2.12]**
Self-employed in the past	0.0774	[7.71]***	-0.0196	[1.41]
Log(Household wealth)	-19.2182	[6.06]***	-5.2953	[1.10]
$[\text{Log(Household wealth)}]^2$	4.5254	[6.74]***	1.0644	[1.05]
$[\text{Log(Household wealth)}]^3$	-0.4964	[7.19]***	-0.1041	[1.00]
$[\text{Log(Household wealth)}]^4$	0.0262	[7.55]***	0.0048	[0.93]
$[\text{Log(Household wealth)}]^5$	-0.0005	[7.79]***	-0.0001	[0.85]
Log(Wage)	-4.4184	[5.30]***	0.4712	[0.43]
$[\text{Log(Wage)}]^2$	0.3555	[4.84]***	-0.0392	[0.40]
$[\text{Log(Wage)}]^3$	-0.0073	[3.38]***	0.0008	[0.26]
Log(Household wealth)*Log(Wage)	-0.0456	[6.52]***	0.0122	[1.26]
Regions:				
East	{*Ref.*}		{*Ref.*}	
North	-0.1372	[14.02]***	0.0220	[1.56]
Central	-0.1052	[10.30]***	0.0133	[0.92]
North-West	-0.0104	[1.00]	-0.0182	[1.27]
South-West	0.0003	[0.04]	-0.0109	[1.08]
South	0.2497	[32.16]***	-0.0165	[1.57]
Inland	-0.1552	[14.26]***	0.0054	[0.35]
Education type:				
Business and Administration	{*Ref.*}		{*Ref.*}	
Humanities & Arts	-0.0862	[9.71]***	-0.0014	[0.11]
Teacher Training & Pedagogy	-0.1890	[11.59]***	0.0091	[0.41]
Social Sciences & Law	-0.2062	[14.25]***	0.0514	[2.57]**
Natural Sciences, Vocational & Technical	-0.1590	[21.42]***	0.0007	[0.07]
Health, Welfare & Sport	-0.1251	[8.38]***	-0.0079	[0.39]
Primary Industries	-0.3570	[18.25]***	-0.0126	[0.48]
Transport & Communications, Safety/Security & other services	-0.1228	[6.81]***	-0.0007	[0.03]
Unspecified broad field of education	-0.1066	[7.86]***	0.0036	[0.19]
Firm size:				
1-10 employees	{*Ref.*}		{*Ref.*}	
10-25 employees	-0.0383	[4.11]***	0.0013	[0.10]
25-100 employees	-0.0347	[4.04]***	0.008	[0.67]
100-500 employees	-0.0096	[1.08]	0.0033	[0.27]
More than 500 employees	0.0316	[3.45]***	0.0043	[0.34]
Industry (ISIC/SN83):				
Activities not adequately defined	0.0764	[3.07]***	0.0073	[0.22]
Agriculture and Hunting	-0.1861	[4.34]***	0.0333	[0.51]
Forestry and logging	-0.0035	[0.06]	0.0499	[0.59]
Fishing	0.1264	[2.65]***	-0.0152	[0.23]
Crude Petrol. & Natural Gas Prod.	0.1976	[11.22]***	0.0223	[1.00]
Metal Ore Mining	-0.1594	[1.50]	0.0206	[0.13]

Table D1 continued in next page

Table D1 continued from last page

	(1)		(2)	
Other Mining	0.0296	[0.61]	-0.0529	[0.81]
Mnf. of Food, Beverages & Tobacco	-0.0653	[3.97]***	0.0238	[1.02]
Textile, Apparel & Leather Ind.	0.0618	[1.40]	0.0033	[0.06]
Mnf. Wood & its Products, Furniture	-0.1226	[5.16]***	0.0269	[0.78]
Mnf. Paper & Products, Print/Publishing	-0.1018	[5.45]***	0.0155	[0.59]
Mnf. Chemicals & Products, Petrol., Coal, Rubber & Plastic Prod.	0.2066	[10.68]***	0.0093	[0.36]
Mnf. Non-Metallic Mineral Prod., except prod. of Petroleum & Coal	0.2929	[10.53]***	0.0015	[0.04]
Basic Metal Industries	-0.0816	[2.87]***	0.0121	[0.29]
Mnf. Fabricated Metal Prod., Machinery & Equipment	-0.0015	[0.13]	0.0174	[1.03]
Other Manufacturing Industries	-0.1196	[2.39]**	0.0167	[0.23]
Electricity, Gas and Steam	-0.0725	[3.88]***	0.0129	[0.49]
Water Works and Supply	-0.2382	[2.59]***	0.0169	[0.13]
Construction	-0.0095	[0.81]	0.016	[1.00]
Wholesale Trade				
Retail Trade	-0.0352	[2.55]**	-0.0049	[0.26]
Restaurants and Hotels	0.0621	[2.28]**	-0.0127	[0.34]
Transport and Storage	-0.0594	[4.96]***	0.0182	[1.09]
Communication	-0.1527	[8.84]***	0.0147	[0.60]
Financial Institutions	0.4972	[22.58]***	0.0387	[1.37]
Insurance	0.9447	[38.22]***	0.0886	[2.90]***
Real estate and Business Services	0.1427	[12.30]***	-0.0054	[0.35]
Public Administration and Defence	-0.0826	[7.10]***	0.0053	[0.33]
Sanitary and Similar Services	-0.1575	[4.65]***	0.0118	[0.24]
Social & Related Community Service	-0.1854	[15.16]***	-0.0137	[0.81]
Recreational and Cultural Services	-0.1202	[4.97]***	-0.0195	[0.58]
Personal and Household Services	-0.1784	[8.73]***	0.013	[0.43]
Constant	46.588	[7.00]***	8.1617	[0.82]
No. of Observations	397,019		162,152	
Pseudo R^2	0.1083		0.0005	
Log-Likelihood	-163,223.9		-95,321.3	

Notes: The table reports estimates from regressions of stock market participation (during the years 1995-1999) for the sample of males aged between 25 and 50 in 1993, not unemployed or self-employed in 1993 or 1994, and not working for a listed company or a subsidiary between 1993 and 1999. Coefficients and z-statistics from probit models are reported. Column 1 reports estimates for the full sample of investors and non-investors, while Column 2 presents estimates for the sample of investors and matched non-investors, using the weights generated as described in the Appendix D. Star levels next to the brackets are levels of significance of the estimates, denoting: * p<0.10, ** p<0.05, *** p<0.01.