Ari Hyytinen – Jukka Lahtonen – Mika Pajarinen

Entrepreneurial optimism and survival

SUOMEN PANKKI
FINLANDS BANK
EUROJÄRJESTELMÄ
EUROSYSTEMET

Bank of Finland Research
Discussion Papers
20 • 2012

Entrepreneurial optimism and survival*

Ari Hyytinen
University of Jyväskylä

Jukka Lahtonen
University of Jyväskylä

Mika Pajarinen
The Research Institute of the Finnish Economy (ETLA)

April 24, 2012

Abstract

This paper uses entrepreneurs' survival expectations around the time of market entry and subsequent venture exits to study entrepreneurial optimism. Using data on a large number of nascent entrepreneurs in the US and start-ups in Finland, we find that new entrepreneurs' survival beliefs are on average optimistic but heterogeneous: Some are excessively optimistic, whereas a small subset holds unbiased beliefs. Entrepreneurial optimism is increasing in the relative (interpersonal) optimism and decreasing in entrepreneurs' level of education and industry experience in both countries. At least in Finland, those holding optimistic views are more likely to transit into entrepreneurship.

JEL: D21, L20
Key words: entrepreneurship, survival, optimism, overestimation

---

*We would like to thank seminar and conference participants in Bergen (NHH, 2010), at the Success-in-Entrepreneurship -conference (CBS, Copenhagen 2010), EARIE 2007 (Valencia), the XXX Annual Meeting of Finnish Economists (Jyväskylä), The Research Institute of the Finnish Economy (ETLA, Helsinki), and the XXIV Summer Meeting of Finnish Economists (Jyväskylä), as well as Gavin Cassar, Thomas Åstebro, Frode Steen, Hans Hvide, William Kerr, Edward Lazear, Markku Maula, Mirjam van Praag, Jaakko Pehkonen, Manju Puri, Anssi Rantala, Petri Rouvinen, Tuomas Takalo, Hannu Tervo, and Otto Toivanen for helpful discussions. This work was initiated in a ETLA/Etlatieto research project on entrepreneurship, financially supported by the National Technology Agency of Finland (Tekes, project 579/31/03) and continued as a part of LIIKE2 project of Tekes and when Hyytinen was visiting the Bank of Finland. Corresponding author: Ari Hyytinen, School of Business and Economics, University of Jyväskylä, P.O. Box 35, 40014 University of Jyväskylä, Finland. Email: ari.t.hyytinen@jyu.fi.

1 Introduction

Economic theory assumes that economic agents are forward-looking and base their major personal and business decisions, such as how much to invest in education, which occupation to choose, and when to retire, on relevant subjective expectations.[1] We study the expectations of entrepreneurs around market entry. Besides being a major life event to most individuals, our motivation to focus on entrepreneurial market entry is that an emerging consensus in the literature appears to be that entrepreneurs are prone to unrealistic optimism. They are, if owner-managers of new businesses for example overestimate the chances of success of their business compared to the actual longevity of the business. If such a positive bias exists, it may both be a driver of entry into entrepreneurship and affect entrepreneurs' post entry decisions, such as investment and hiring decisions.

Many papers draw and build on the conclusion that entrepreneurs are optimistic.[2] However, somewhat surprisingly, direct field evidence for (absolute) entrepreneurial optimism (overestimation; positive expectational bias) at the level of individual decision-makers or ventures is quite limited.[3] Typical cited evidence that uses field data either cannot relate entrepreneurs' expectations to

[1] The most obvious empirical counterpart of the subjective expectations are the self-reports of subjective probabilities that are increasingly elicited in surveys and laboratory experiments (Manski 2004). Yet, despite many recent contributions (see, e.g., Basset and Lumsdaine 2001, Clark and Friesen 2009, Manski and Molinari 2010), the literature on the accuracy of and potential biases in probabilistic expectations is still limited in scope.

[2] Recent economic models that build on the notion of optimism include de Meza and Southey (1996), Manove and Padilla (1999), Bernardo and Welch (2001), Hyytinen (2003), Brocas and Carillo (2004), Coval and Thakor (2005), and Landier and Thesmar (2009). Van den Steen (2004), Santo-Pinto and Sobel (2005) and Brunnermeier and Parker (2005) are examples of more general models of (optimally) biased expectations. In management and organizational science, a number of papers, such as Hayward, Shepherd and Griffin (2006) and Dushnitsky (2010), refer to overly optimistic entrepreneurs.

[3] Building on an established but still growing psychological literature, Moore and Healy (2008) distinguish three types of overconfidence: Overestimation, overplacement and overprecision. The first of these refers to a miscalibrated forecast or expectation relative to the objective likelihood of an event, the second to the better-than-average -effect (i.e. interpersonal optimism), and the third to the tendency of people to be overly confident on how accurate their forecasts or expectations are.

outcomes at all (e.g., Fraser and Greene 2006), compares entrepreneurs' predictions only to aggregate outcomes (e.g., Coopers, Woo and Dunkelberg 1988, Koellinger, Minniti and Schade 2007), focuses on overprecision in estimation or overplacement (e.g., Forbes 2005, and Ucbasaran, Westhead, Wright and Flores 2010), or measures entrepreneurial optimism indirectly using data on respondents' expectations about general economic outcomes or life events (e.g., Arabsheibani, de Meza, Maloney and Pearson 2000, Puri and Robinson 2007, 2009).[4]

The two prior studies which directly measure entrepreneurial optimism and which our analysis complements are Landier and Thesman (2009) and Cassar (2010).[5] The study of Landier and Thesman focuses on the effects of entrepreneurial optimism on financial contracting. Using the SINE survey of French entrepreneurs and tax data from the 1990s, they document that entrepreneurs' optimism about the future development of their firm and hiring needs covaries with their reliance on short-term debt. Landier and Thesman also show that having less education and more industry experience are associated with less optimistic expectations. Cassar, in turn, uses the US Panel Study of Entrepreneurial Dynamics (PSED) to explore the determinants of new entrepreneurs' optimism about their future employment needs and about the likelihood that their nascent venture eventually becomes operative. Cassar's results show that in the PSED data, it is surprisingly difficult to explain entrepreneurs' optimism with their and their ventures' observable characteristics. For example, and in contrast to Landier and Thesman, he does not find statistically significant associations between entrepreneurs' optimism and their education and industry

[4] Indirect evidence from economics is also somewhat mixed. The available analyses support neither the view that positive biases in agents' absolute forecasts of (economic) outcomes are universal nor that they would be systematic across different decision contexts or forecasting environments (Clark and Friesen 2009, Grieco and Hogart 2009).

[5] We should also mention the study by Arabsheibani, de Meza, Maloney and Pearson (2000), which uses the British Household Panel Survey from 1990-1996 to document that self-employed are more optimistic than others.

experience.

In this paper, we study the survival expectations of entrepreneurs around market entry using a large number of recently founded start-ups in Finland as well as the nascent entrepreneurs in the US PSED data. These two data sets allow us to construct a (comparable) measure of entrepreneurial optimism as the difference between the objective probability that the venture of an entrepreneur is not in business after a certain period of time and the corresponding subjective probability elicited from the entrepreneurs around the time of entry.

Our analysis confirms some of the results presented earlier in the literature and generates new insights. First, we find - consistent with Landier and Thesman (2009) and Cassar (2010) - that new entrepreneurs hold substantially optimistic survival beliefs both in the US and in Finland. Second, we show that though the underlying data sets are somewhat different, entrepreneurial optimism displays similar patterns in the two countries. For example, we provide evidence on the quantitative importance of heterogeneity in the degree of entrepreneurial optimism: In both countries, many new entrepreneurs are excessively optimistic, but a rather small subset holds unbiased survival beliefs.[6] This documented heterogeneity bears on the psychological and economics literature on dispositional optimism, which argues that heterogeneity matters and has normative implications: A moderate degree of optimism may lead to economic behaviour that is more likely to result in desired outcomes, but excessive optimism can lead to suboptimal decisions (see, e.g., Rabin 1998, and Puri and Robinson 2007).[7] In the context of market entry, excessive optimism is of direct policy-interest, as it questions the need for public policies that generically and indiscriminately boost entry.

[6] Note that heterogeneity refers here to the cross-sectional variability of entrepreneurs' forecast *errors*, not just to the variability of their subjective beliefs.

[7] The evidence provided by Puri and Robinson (2007) squares nicely with this view. Their results suggest that there is considerable heterogeneity in the degree of optimism across the households and that the heterogeneity covaries with the prudence of financial behaviour.

Unlike the preceding studies, we are able to document that both in Finland and in the US, (absolute) entrepreneurial optimism is positively correlated with the relative (better-than-average) optimism. This finding is directly linked to the laboratory experiments that focus on market entry games with biased beliefs. Papers in this strand include Camerer and Lovallo (1999), Coelho, de Meza and Reyneirs (2004), and Bolger, Pulford and Colman (2008). Camerer and Lovallo suggest, for example, that when success is contingent partly on skill, entry is excessive due to reference group neglect. This neglect means that the entrepreneurs who self-select to enter a market are overly optimistic about their own level of skills relative to the level of skills of the other entrants. Taking this finding as their starting point, Bolger et al. make a distinction between interpersonal comparison of abilities and the overestimation of one's abilities in absolute terms. Bolger et al. find that when skills matter, absolute (rather than relative) overconfidence drives excessive entry decisions. We show that the two are interrelated in the field. If relative optimism does not objectively reflect an entrepreneur's interpersonal advantage, positive correlation between the two is what we ought to expect.

We also find that having more industry experience is associated with less optimistic expectations both in the Finnish and US data. This finding is consistent with the conventional wisdom that experience reduces nonstandard behavior in markets (e.g., List 2003; see also DellaVigna 2009, p. 365) and squares nicely with the findings of Landier and Thesman (2009). The role of education is, however, less clear, as we find that having *more* education is associated with *less* optimistic expectations. This relation is consistent with Arabsheibani, de Meza, Maloney and Pearson (2000), who find that higher education is associated with smaller optimism in the UK, and sharpens Cassar's (2010) findings for the US, which are a bit mixed in this regard. However, the relation that we find is in contrast to the finding of Landier and Thesman for France, who argue that the highly educated may have better outside options and therefore become entrepreneurs only when they are exceptionally optimistic.

While the fact that new entrepreneurs hold positively biased expectations of their survival probability suggest that there may be too much entry, the question of whether optimism around market entry has real behavioural consequences remains. We therefore use another Finnish data set to complement our main analysis. These data allow us to observe individuals' labour market transitions and to construct a measure of optimism similar to that used by Arabsheibani, de Meza, Maloney and Pearson (2000), who make use of economic expectations of individuals and the corresponding realizations. Using these data, we take a look at whether optimism is positively correlated with *transitions* into entrepreneurship.[8] It indeed is. We find, in particular, that those who are more optimistic are more likely to become entrepreneurs than their non-optimistic counterparts.

The plan of the paper is as follows. Section 2 presents our main analysis. It focuses on entrepreneurial optimism about venture survival around the time of market entry and studies its level, heterogeneity and correlates using data both from the US and Finland. Section 3 uses the complementary Finnish data to examine whether optimism is associated with individuals' labour market transitions. Section 4 offers concluding remarks.

2 Entrepreneurial optimism about survival

After describing the US and Finnish data sets available for our empirical analysis, this section addresses two main questions: First, are entrepreneurs optimistic

[8] Arabsheibani, de Meza, Maloney and Pearson (2000) do not study whether optimism is associated with transitions into self-employment.

about the survival of their start-ups? If they are, are all entrepreneurs equally optimistic or is there heterogeneity in the degree of their optimism? Second, what observable characteristics predict optimism?

2.1 Data sources

We use Finnish and US data sets in this section. The first merges two nationwide cross-sectional surveys of Finnish entrepreneurs (owner-managers), who at the time of the surveys, had just started to run their start-ups. The second is based on a nationally-representative sample of nascent entrepreneurs in the US, called the Panel Study of Entrepreneurial Dynamics (PSED) and used earlier by Cassar (2010).[9]

2.1.1 Finnish data set

The two Finnish surveys targeted new start-ups and were based on computer-aided telephone interviews.[10] The target start-up population of the first of them, the 2003 survey, were those 2207 firms which had been granted a new business identity code by the Finnish Trade Register in October 2003. The objective of the survey was to target these firms, but due to the confidentiality of contact information of the smallest businesses (annual turnover $<$ 8500 euros) and non-response from the registered phone number (despite numerous attempts), 870 (39%) firms were eventually reached for an interview. The target start-up population of the second survey, the 2005 survey, were those 13,477 new firms that had entered Statistic Finland's "Enterprise openings and closures" -database in the first half of 2005 and that satisfied certain minimal criteria for being a new business. Approximately every fourth (23%) firm was randomly chosen to be

[9] See http://www.psed.isr.umich.edu/psed/home for the documentation of the data.

[10] Rouvinen and Yla-Anttila (2004) and Pajarinen, Rouvinen and Yla-Anttila (2006) describe the 2003 and 2005 surveys in detail.

included in the survey, resulting in a target sample of 3042 firms.[11] Out of this sample, 1,888 firms (62 %) were eventually reached for an interview, because the rest either had no active or no non-confidential contact information. While we cannot conclusively rule out selection problems, it is not obvious why these two (technical) reasons for non-response would cause systematic biases in our empirical results for Finland.

At the time of the 2003 survey, firms were approximately four months old, whereas at the time of the 2005 survey, the firms were, on average, six months old. The number of completed interviews in the 2003 survey was 393 and in the 2005 survey, 616. Due to non-response to certain key questions, we have 891 observations in our final, merged estimating sample.

All variables except for the indicator of firm exit come from the surveys. The indicator for firm exit is derived from the register data of Statistics Finland, using data three years after the initial surveys. We discuss the exact measurement of exits in this data below.

2.1.2 US data set

The objective of PSED was to identify and interview nascent entrepreneurs who are in the process of founding a new venture (see also Cassar 2010). In 1998-2000, an initial screening of over 64,000 individuals from the US mainland resulted in a sample of 3,592 respondents who were considered eligible for "nascent entrepreneur interview". A random sub-sample of 1,164 was selected for a further interview. Out of them, 830 respondents was after additional screening eventually selected for three follow-up interviews. These respondents were at the time trying to start a new business. Being eligible for the survey required that the new business was not able to support a positive cash flow yet.

[11] The survey also had a component that overweighted limited liability firms. We work with the random sample.

Our sample is based on the 830 nascent entrepreneurs that were eligible for the three follow-up surveys. It excludes those who were attempting to start business on another persons' behalf, who did not report the ex-post status of the venture within five years of the original interview or who did not reply to the survey questions that we use to construct our key explanatory variables. Our final estimating sample consists of 347 nascent entrepreneurs.[12]

All variables in the US data are from the PSED surveys, including the indicator for venture exit.

2.2 Variable definitions and descriptive statistics

2.2.1 Survival expectations and outcomes

In the Finnish data set, our primary measure of the subjective probability that a start-up goes out of business within a three-year period, IFAIL, is based on question: "*Please estimate the probability that your venture exits and is not in business three years from now?*" The indicator for firm exit, D_EXIT, coming from the register data of Statistics Finland, indicates the status (still-in-business vs. not-in-business) of the surveyed start-ups approximately 3 years after the surveys. This timing matches with the duration to which the initial survey question on the probability of survival of the start-ups applies. Statistics Finland determines the status of firms continuously, using information from various subregisters, such as business, tax, and employment registers. It is worth mentioning that Statistics Finland has access to plant-level register data, which reduces the risk of misclassification of exits due to, e.g., changes in the business identity codes that are related to M&As.

In the US data, the subjective survival probability is elicited using the ques-

[12] We acknowledge that there may be selection in and out of the US PSED sample. Most of the attrition out of the PSED sample is due to the fact that some of the entrepreneurs that responded to the initial survey did not participate in the follow-up surveys. It is not obvious why the patterns of non-response would be systematic enough to lead to biased inference.

tion "*On a scale of zero to one hundred, what is the likelihood that this business will be operating five years from now, regardless of who owns and operates the firm?*". We construct IFAIL as one minus the answer to this question (scaled by 100). The indicator for firm exit, D_EXIT, is based on the status of the start up five years after the initial interview, as reported by the entrepreneur(s) in the follow up surveys conducted in three waves, at a mean of 14, 33, and 56 months following the initial interview. In particular, we set D_EXIT to one if the respondent reported at some stage along the three follow up interviews that the nascent activity or operating business is no longer worked by anyone and it is set to zero if the respondent reported in the last interview that the business is operational.[13]

Figure 1 summarizes the data sets used in this section and the timing of measurement. The Finnish data refer to entrepreneurial optimism conditional on entry (i.e., post-entry optimism), whereas the US data is about entrepreneurial optimism before entry (i.e., pre-entry optimism). Together with the difference in the duration of the period (3 vs. 5 years) to which the beliefs apply, this means that the two data sets are complementary but not directly comparable.

[Insert Figure 1 about here]

Before proceeding, we note that it is, of course, easy to criticize our survey-based expectation measures, as eliciting accurate probabilistic expectations using standard survey instruments is not easy (Bassett and Lumsdaine 2001). Nor is there full agreement on how it should be done. On the one hand, individual respondents may have weak incentives to provide reliable forecasts, as they are

[13]We ought to mention we are not able to replicate Cassar's exit-indicator perfectly despite our best efforts. Following as closely as we can his definitions and sample restrictions, we find 142 (208) ventures for which we would set D_EXIT to zero (one). In his largest sample, the corresponding numbers are 185 and 201 (Cassar, 2010, Table 1, p. 828). If we rerun our main regressions using this (best-effort replicating) sample, the size of the sample reduces to 276 due to missing observations of the explanatory variables. We can nevertheless report that the results of these regressions are similar to the findings that we get when we use our definitions for exit to construct the US sample.

not rewarded for the accuracy of their predictions. Knowing that their accuracy is not challenged, the respondents do not necessarily deliberate the question carefully and may even practice some sort of window-dressing so as "to look or sound good" when they are interviewed. On the other hand, it is often argued that people may have difficulties in understanding and formulating probabilities even if they are motivated to do so.

We have four responses to these concerns. First, a large number of well-known large-sample surveys already use probabilistic formats to elicit expectations and it seems that non-response to probabilistic questions or systematic biases (or one-sided window-dressing) are *not* general features of these data (Bassett and Lumsdaine 2001, Manski 2004).[14] Second, the recent literature does not unequivocally support the view that rewarding for the accuracy of (probabilistic) predictions matters. Clark and Friesen (2009) document, for example, that the probabilistic forecasts were *not* more accurate when their accuracy was rewarded. Third, a unique feature of the Finnish data set is that it allows us to measure the time it took (in seconds) for a respondent to answer the question that elicited the expectation. If it takes less time to give a sloppy window-dressing response than a carefully deliberated and honest one, we should find that the response time is negatively correlated with either the *stated* likelihood of survival or the accuracy of the survival belief. They are *not*.[15] Fourth, as explained above, the probabilities were elicited in different ways in the US and Finnish surveys. The US survey elicited probabilistic estimates as numbers between one and 100, which at least some scholars seem to prefer (e.g., Viscusi 1990, Clark and Friesen 2009), whereas the Finnish surveys directly asked for the probabilities of interest. While it is unclear whether the

[14] Such surveys include but is not limited to Health and Retirement Study (US), the Survey of Economic Expectations (US), Michigan Survey of Consumers (US), Survey of Household Income and Wealth (Italy), and VSB Panel Survey (Holland).

[15] For example, the pairwise correlation between the response time and the stated likelihood of survival is close to zero (-0.024) and not significant statistically.

method of elicitation results in a particular type of bias in the data, it is of interest to note already here that the US data explored suggest a greater degree of optimism (see below).

2.2.2 Descriptive analysis

Table 1 takes a first look at the distribution of IFAIL in the two data sets. It shows that in Finland, about 90% of the respondents think that the likelihood of exit of their new venture is 40% or less in three years time. In the US, 83% of the nascent entrepreneurs think that the likelihood that their business is not operating five years from now is less than 40%.

The mean of IFAIL is 17% in the US and 12% in Finland. These means are much lower than the means of D_EXIT, which are 65% (US) and 34% (Finland). The difference between these numbers suggest that both in Finland and in the US, entrepreneurs are optimistic on average. The difference is about 48 and 22 percentage points, respectively, in the US and Finland. Besides heterogeneity in national characteristics, the difference in the degree of optimism that the above numbers imply can be either because some US entrepreneurs fail to enter or because conditional on entry, they are at risk of failing for a longer period than their Finnish counterparts.

[Insert Table 1 about here]

We can also read from Table 1 that the pairwise correlation between IFAIL and D_EXIT is 0.21 and highly significant (p-value <0.01) in Finland, but only 0.08 and insignificant in the US. Albeit low, these positive correlations indicate that those entrepreneurs who think that they won't exit are, on average, less likely to exit.

Figure 2 describes the mean of D_EXIT, conditional on IFAIL being in one of the "0", "1-10", "11-20", "21-40", "41-100" percentage categories. In each category, the two leftmost bars display the mean of D_EXIT and the two rightmost bars the mean of IFAIL in the category. Note that in category "0", the mean of IFAIL is, by definition, zero. If the subjective probabilities were well calibrated, we would expect the height of the D_EXIT bars to roughly match that of the IFAIL bars. This is clearly not the case.

The figure suggests, in particular, that there is considerably heterogeneity in the degree of optimism. It shows that more than 60% of the businesses of those US nascent entrepreneurs who think that their business is operative with probability one five years from now (i.e., those in category "0") are not, in fact, operative five years after the survey. The corresponding percentage is much smaller in the Finnish sample but still nearly 30%. In both countries, the differences in the degree of miscalibration between the categories are clear, indicating that the magnitude of the expectation bias varies with the subjective probability.[16]

[Insert Figure 2 about here]

Table A1 and A2 in Appendix 1 provide the definitions of the explanatory variables used in the regressions and give their descriptive statistics both for the Finnish and US samples.

2.3 Empirical analysis

2.3.1 Measuring bias in survival beliefs

Our aim is to develop a measure for optimism, i.e., for the overestimation of the likelihood of an event, in the particular context of market entry. Our primary measure for the bias is the difference between the objective probability that

[16] The figure suggests, however, that the subjective probability expectations and exit outcomes are not independent, as the higher is an entrepreneur's IFAIL, the more likely that his firm is no longer in business after three years. The relation is, however, clearly weaker for the US than Finnish entrepreneurs. This difference is consistent with the cross-country difference in the pairwise correlations that we documented above.

the venture of an entrepreneur exits (is no longer in business) within a certain period of time and the subjective probability of the same binary event.

To be a bit more formal, let e_i denote how erroneous (biased) the exit belief of entrepreneur i is. We define this bias in the survival beliefs to be equal to the difference between the objective probability that the venture of entrepreneur i exits within a certain period of time and the subjective probability of the same binary event, i.e.,

$$e_i \equiv \pi_i(p_i) - p_i \tag{1}$$

where $\pi_i(p_i)$ denotes the objective exit probability and p_i the subjective probability.

As we explained above, we have data on the subjective probability, as this probability has been elicited directly from the entrepreneurs in the surveys. However, we do not observe the objective probability. Instead, we observe whether the venture of an entrepreneur actually exits or not, denoted here Y_i for brevity. We postulate the following probability model for Y_i:

$$\pi_i(p_i) = \Pr\left[Y_i = 1 \,|\mathbf{X}_i, p_i\,\right] = F(\mathbf{X}_i'\beta + \gamma g(p_i)), \tag{2}$$

where $\mathbf{X}_i$ is a vector of covariates that help predicting exit, $g(p_i)$ is some (known, possibly) non-linear function of p_i, $\theta \equiv (\beta, \gamma)$ is a vector of model parameters and the distribution function, $F(\cdot)$, is for the error term that captures all those firm and entrepreneur characteristics that affect the likelihood of exit that we (as econometricians) are not able to observe.

We make note of three features of this model:

First, assuming that the data, $(Y_i, \mathbf{X}_i, p_i)$ for $i = 1, 2, ..., N$, are *i.i.d.* and have been obtained by random sampling from some population, ML-estimation gives estimates $\widehat{\theta} = (\widehat{\beta}, \widehat{\gamma})$ of the corresponding unknown population parameters. They allow us to compute predicted exit probabilities, $\widehat{\pi}_i(p_i) = F(\mathbf{X}_i'\widehat{\beta} + \widehat{\gamma}F^{-1}(p_i))$ for $i = 1, 2, ..., N$. This means that we can replace unobserved e_i's with observed $\widehat{e}_i$'s, i.e., with

$$\widehat{e}_i = \widehat{\pi}_i(p_i) - p_i. \tag{3}$$

Second, $\widehat{e}_i$ is consistent for e_i, which justifies using $\widehat{e}_i$ as a measure for the bias in entrepreneurs' beliefs. The observed measure, $\widehat{e}_i$, may differ from unobserved e_i, because $e_i - \widehat{e}_i = \pi_i(p_i) - \widehat{\pi}_i(p_i) = \widehat{v}_i$, where $\widehat{v}_i$ is an estimation error. However, provided that the model is correctly specified, the ML-estimator, $\widehat{\theta} = (\widehat{\beta}, \widehat{\gamma})$, is consistent for $\theta = (\beta, \gamma)$. In large samples we thus have $F(\mathbf{X}_i'\widehat{\beta} + \widehat{\gamma}F^{-1}(p_i)) \xrightarrow{p} F(\mathbf{X}_i'\beta + \gamma F^{-1}(p_i))$ and $\widehat{\pi}_i(p_i) \xrightarrow{p} \pi_i(p_i)$ by Slutsky's theorem. This implies that $\widehat{e}_i \xrightarrow{p} e_i$. We therefore have a measure of optimism that is in large samples asymptotically unbiased for each entrepreneur.[17]

Third, we allow $\pi_i(p_i)$ to be an explicit function of the subjective probability for two reasons. On the one hand, the subjective probability is a determinant of the effort that an entrepreneur is willing to exert to make his venture successful. On the other hand, we want to allow for the possibility that subjective probabilities have cross-sectional predictive power of venture exits. In particular, we let $g(p_i)$ be monotonically increasing in p_i and impose $g(p_i) = F^{-1}(p_i)$ in the empirical application. This implies that $d\pi_i(p_i)/dp_i = \gamma$ and that the model nests the possibility that the subjective exit probabilities are positively (but not one-to-one) related to the model-based likelihood of exit (i.e., $\gamma \in (0, 1)$).[18]

An alternative way to measure optimism is to compute $\widetilde{e}_i = Y_i - p_i$, which

[17] We acknowledge that it is easy to construct examples which demonstrate the finite sample bias of a ML-estimator. In our application, the sample size is large enough to allow for asymptotic arguments.

[18] Model (2) has also the following properties: Conditional on $\gamma = 1$, $\mathbf{X}_i'\beta$ measures the bias of beliefs for a subject with characteristics $\mathbf{X}_i$ because then $e_i = 0 \Leftrightarrow \mathbf{X}_i'\beta = 0$. However, if $\gamma \neq 1$ is allowed, then $\mathbf{X}_i'\beta$ is a measure of local bias at $p_i = 0.5$. The model also allows for perfect calibration of beliefs. If $\gamma = 1$ and $\beta = \mathbf{0}$, $F\left[F^{-1}(p_i)\right] - p_i = e_i = 0$ for all i.

has been used in Cassar (2010) and which is close in spirit to the method used by Das and van Soest (1997, 1999) to study the accuracy of households' income expectations. It is worth pointing out, however, that in light of (1), Y_i is an imperfect measure of $\pi_i(p_i)$. Though the difference ($\widetilde{e}_i$) can be used to measure whether the entrepreneurs are optimistic *on average*, it is not a good measure for individual-level outcomes; clearly, $\widetilde{e}_i$ can be zero only for those exiting (surviving) ventures who predict failure (success) with probability one. This means, for example, that the measure is not well suited for interpersonal comparisons or studying the quantitative importance of heterogeneity in optimism.

Moreover, $\widetilde{e}_i$ is "asymptotically noisier" for e_i than $\widehat{e}_i$. To see why, let $Y_i = \pi_i(p_i)+\omega_i$, where ω_i is a regression error. Comparing $e_i-\widehat{e}_i = \widehat{v}_i$ with $e_i-\widetilde{e}_i = \omega_i$ we observe that the former converges to zero as the size of the sample increases, whereas the latter does not.

The above arguments suggest that (3) may have some econometric advantages over $\widetilde{e}_i$. However, to err on the safe side and to illustrate the robustness of our findings, we also use $\widetilde{e}_i$ as an alternative measure in the relevant parts of what follows.

2.3.2 Mean bias and heterogeneity

Are entrepreneurs optimistic about the survival of their start-ups? How much variation there is in entrepreneurial optimism? To address these questions, we have estimated (2) as a Logit, Probit and Cloglog -model for our baseline analysis. The explanatory variables include $F^{-1}(p_i)$, ifail_notzero (= a dummy that is set to one if $p_i > 0$ and is zero otherwise), as well as additional 14 regressors for the US and 20 regressors for Finland. We do not display the coefficient estimates of these variables here in the main text, because our main interest is in the predicted probabilities of these models.[19] Instead, we note that

[19] The estimated models are displayed in Appendix 2, Table A3 and A4.

in all models, the explanatory variables are jointly significant and present the estimated density functions of the probability difference, $\widehat{e}_i$.[20]

The densities can be found in Figure 3 (US) and 4 (Finland). Two things stand out. First, the distributions seem to be well behaved, i.e., they are relatively smooth and symmetric. Second, most of the mass of the distributions is at the right side of zero. This finding is consistent with the view that entrepreneurs are optimistic on average. The mean of the US distribution is 0.48 whereas that of the Finnish data is 0.22.

[Insert Figure 3 and 4 about here]

It is also worth noting that the differences in $\widehat{e}_i$'s generated from the Logit, Probit and Cloglog -models are minor, if not negligible. The pairwise correlations of the predicted probabilities between the three models confirm this perception as they are always higher than 0.988 and significant at better than the 1% level for both countries. We therefore focus in what follows on $\widehat{e}_i$'s that have been computed from the estimated Logit models.[21] Though this may at first sound to be a restrictive choice, we show later that our main results are robust to using a semi-parametric binary model and the alternative measure of optimism ($\widetilde{e}_i$).

Besides documenting the cross-sectional mean bias in the survival beliefs of entrepreneurs, we can examine the heterogeneity in the degree of entrepreneur-

[20] We also tested whether the subjective probabilities have cross-sectional predictive power for venture exits, conditional on the other regressors. Consistent with the higher pairwise correlation in Finland, the joint tests for the significance of $F^{-1}(p_i)$ and ifail_notzero indicate that in the Finnish data set, the subjective probabilities have statistically significant predictive power (p-value < 0.01, Logit model), whereas in the US data, they have not (p-value 0.2806, Logit model). In the Finnish data, the marginal effect of p_i is around 0.3 (p-value < 0.01), conditional on $p_i > 0$. This suggests that a 1 percentage point increase in p_i increases the predicted probability of venture exit by 0.3 percentage points. For the US, the corresponding marginal effect is smaller and not significant.

[21] We apply no weights to the PSED data, because due to attrition, it is unclear how appropriate doing so would be. Our results are not, however, sensitive to this choice. Using the estimated Logit models with and without sample weights we find that the pairwise correlation between the unweighted and weighted $\widehat{\pi}(p_i)$ is 0.983 (p-value$<$0.01). Moreover, the pairwise correlation between the unweighted and weighted $\widehat{e}_i$ is 0.990 (p-value$<$0.01).

ial optimism: Given that no entrepreneur knows her probability of survival for sure in advance (and given that they are on average optimistic), are all entrepreneurs *equally* optimistic? Or, is there heterogeneity in the degree of their optimism? Can we, for example, find some entrepreneurs who hold unbiased views? Do some entrepreneurs appear to be much more optimistic than others? Is heterogeneity quantitatively important?

Quantiles of e_i provide a view on heterogeneity and they have a direct link to the mean bias, as $E[e_i] = \int G^{-1}(\tau)d\tau$, where $G(e_i)$ refers to the c.d.f. of e_i and $G^{-1}(\tau)$ denotes the τ^{th} quantile of e_i. If there is negligible heterogeneity, $G^{-1}(\tau)$ ought not to vary with τ. We provide a characterization of the distribution of e_i and its quantiles using standard quantile regression methods (Koenker 2005).

Table 2 shows the estimated quantiles of $\widehat{e}_i$. The fifth decile (the median) of the difference between the objective probability that the venture of an entrepreneur exits and the corresponding subjective probability is 50 percentage points in the US and 20 percentage points in Finland. These medians are estimated relatively accurately, as their 95% confidence intervals are [0.48, 0.54] and [0.20, 0.23], respectively. The table also shows that the US entrepreneurs at the ninth decile of the optimism distribution are 1.5 times more optimistic than those at the median, indicating substantial heterogeneity. In Finland, the corresponding 90-50 ratio is of similar magnitude, about 2.1.

[Insert Table 2 about here]

We can provide complementary evidence on heterogeneity of optimism by decomposing the mean bias to the separate contributions of optimists (i.e., those with $e_i > 0$), realists (i.e., those with $e_i \approx 0$) and pessimists (i.e., those with $e_i < 0$). By the law of iterated expectations, the mean bias in the beliefs is the mean for the optimists times their proportion in the population plus the mean for the realists times their proportion in the population plus the mean for the pessimists times their proportion in the population. To determine the three proportions, we compute a t-test for H_0: $e_i = 0$ for each i. If the null hypothesis is rejected for i, $e_i > 0$ ($0 <$) is consistent with i being optimistic (pessimistic). If the null hypothesis is not rejected for entrepreneur i, he is regarded as a realist. We use a *multiple testing procedure* (MTP) due to Benjamini and Hochberg (1995) to take into account the fact that we test here a large number of hypothesis simultaneously.[22] MTPs control for the expected proportion of falsely rejected null hypothesis, i.e., false discovery rate (FDR), at a desired level.[23] Besides the original Benjamini and Hochberg -procedure, we control for the FDR using the more recent procedure of Benjamini, Krieger, and Yekutieli (2006).

The MTPs amount to i) testing H_0: $e_i = 0$ for each i, ii) computing and ordering the associated p-values in an increasing order and iii) using a sequence of critical values to identify entrepreneurs for whom the null hypothesis is rejected. The critical values are adjusted so that they control for FDR at 5 %. Entrepreneur i is then inferred to hold optimistic (pessimistic) beliefs if the prediction error is positive (negative) and if he belongs to the subsample of entrepreneurs for whom the null hypothesis is rejected.

Table 3 displays the results of the MTPs and the decomposition of the data to optimists, realists and pessimists.[24] It shows that the null hypothesis of realistic exit beliefs is rejected for 87% and 67% of the entrepreneurs in the US and

[22] Standard (independently implemented) t-tests are inappropriate, because they would lead to (too) many false positives. Controlling for familywise error rate is also problematic, because it focuses on the probability of making one or more false positives. This is unduly conservative when the number of tests is large and results in (too) many false negatives.

[23] FDR refers to the fraction of false rejections among all the hypotheses that are rejected. In our application, it refers to the expected classification error among the entrepreneurs that we label as not holding "rational beliefs". A control for FDR is needed, because testing of more than one hypothesis at a time means that the larger the number of tests, the more likely it is that some true null hypotheses get rejected. Control of FDR leads to much higher power than control of FWER. See, e.g., Benjamini and Hochberg (1995) for further discussion.

[24] These numbers are based on the procedure of Benjamini and Hochberg (1995). The results from the procedure of Benjamini et al. (2006) are qualitatively similar but, consistent with the quantile regression analysis, suggest a higher fraction of optimists for both the US and Finland.

Finnish samples, respectively. These findings are broadly consistent with the results of Table 2, where, however, also the lowest deciles of the distribution of $\widehat{e}_i$ are positive and significantly different from zero. The results of the MTPs nevertheless complement the earlier analysis and are consistent with the existence of heterogeneity: The result suggests that there is a subset of entrepreneurs who do not have optimistic survival beliefs. The MTPs show, in addition, that nearly all of the entrepreneurs for whom the null of realistic beliefs is rejected, have optimistic views of venture survival, i.e., $\widehat{e}_i > 0$. This is an important piece of information that a simple graphical analysis (see Figure 3-4) does not reveal: There are entrepreneurs who *appear* to make pessimistic forecast errors, but which are mostly insignificant statistically.[25]

[Insert Table 3 about here]

Taken together, Figure 2-4 and Table 2 and 3 provide clear evidence for the heterogeneity of entrepreneurial optimism at the time of market entry. Entrepreneurs are on average optimistic about their chances of survival, but the average masks a great deal of variation in how prudent the survival beliefs are. It seems, in particular, that a number of new entrepreneurs are excessively optimistic and that a rather small subset holds unbiased beliefs. In light of the literature on dispositional optimism (Rabin 1998, and Puri and Robinson 2007), excessive optimism around market entry is suggestive of suboptimal entry decisions by some (but not by all) new entrepreneurs. This finding, in turn, bears, e.g., on the appropriate design of industrial policy (see the concluding section for further discussion).

[25] An advantage of this approach is that it allows us to compute the mean optimism bias conditional on belonging to the group of optimists. They are 0.55 and 0.31 in the US and Finnish samples, respectively.

2.3.3 Correlates of survival optimism

What observable characteristics predict optimism? To address this question we explore the correlates of optimism using standard linear regression and $\widehat{e}_i$ as the dependent variable.[26] Regressions that use the alternative measure for optimism, $(Y_i - p)$, as the dependent variable yield similar results; see our robustness tests below.

Table 4 displays the results for the regressions in which the vector of explanatory variables, $\mathbf{Z}_i$, includes OTHERSFAIL (= each entrepreneur's perceived exit likelihood of other similar firms), dif_OFAIL_IFAIL (= OTHERSFAIL - IFAIL), D_OTHERSFAIL (= 1 if there is no response to OTHERSFAIL, and = 0 otherwise), EXPERIENCE (= sum of the respondent's work and entrepreneurial experience in the field (industry) of the start-up in years), ACADEMIC and COM_COLLEGE (= 1, respectively, if respondent has a university degree or a type of community college degree, and = 0 otherwise), COMP_ANAL (=1 if respondent prepared systematic financial plans (US) or gathered information about competition (Finnish) prior to entry, = 0 otherwise), and, in the Finnish sample only, D_2003 (=1 if start-up in 2003, = 0 if in 2005).

We include dif_OFAIL_IFAIL, as it can be regarded as a measure of relative (interpersonal) optimism and overplacement. The higher this measure, the more positive an entrepreneur is about the survival of his own venture relative to that of the other ventures. If a high value of this measure reflects an entrepreneur's true information about his interpersonal advantages relative to other entrepreneurs, dif_OFAIL_IFAIL should obtain a negative coefficient. In contrast, if it reflects unwarranted interpersonal comparisons, it should be

[26] The focus on the standard linear regression can be motivated by the notion of conditional expectation function (CEF). For e_i, the CEF is $E[e_i | \mathbf{Z}_i]$, the expectation of e_i conditional on some covariates $\mathbf{Z}_i$. The CEF has a direct link to the mean bias, as $E[e_i] = E\{E[e_i | \mathbf{Z}_i]\}$, where the outer expectation is taken w.r.t. the distribution of $\mathbf{Z}_i$. We use the standard regression to model $E[e_i | \mathbf{Z}_i]$, because it provides the best (minimum mean squared error) linear approximation of the CEF even if $E[e_i | \mathbf{Z}_i]$ is nonlinear.

positively correlated with $\widehat{e}_i$.

To control for "base rate", i.e., the prior probability unconditioned on featural or idiosyncratic evidence, we include OTHERSFAIL. Because some of the respondents do not answer to the OTHERSFAIL question, we replace the missing responses by the mean value (computed over the sub-sample for which the measure is available) and include D_OTHERSFAIL to control for the non-response.

We use EXPERIENCE as an additional explanatory variable, because the prior literature suggests that it might be related to the process that generates biased beliefs (see, e.g., List 2003 and Fraser and Greene 2006). List (2003), for example, shows that marketlike experience is negatively associated with the size of a known and well-documented behavioral anomaly, the endowment effect. The findings of Landier and Thesman (2009) also support the inclusion of EXPERIENCE, as they find that that having more industry experience is associated with less optimistic expectations. Inspired by these prior analyses, we examine whether industry (market) experience moderates entrepreneurial optimism.[27]

We also analyze whether education predicts optimism in the context of market entry. Our motivation to do so is that the earlier results are a bit mixed: Cassar (2010) finds insignificant effects for the US, whereas the finding of Landier and Thesman (2009) suggests that the highly educated entrepreneurs hold more optimistic expectations in France. Arabsheibani, de Meza, Maloney and Pearson (2000) show, in turn, that having better education is associated with less optimism in the UK. We explore whether the same holds in the Finnish and US contexts by including ACADEMIC and COM_COLLEGE -dummies.

Finally, COMP_ANAL is included for two reasons. First, Cassar (2010)

[27] Because for some of the U.S. respondents we cannot measure EXPERIENCE, we replace the missing responses by the mean value (computed over the sub-sample for which the measure is available) and include D_EXPERIENCE to control for the non-response.

argues that those entrepreneurs who use an inside (i.e., subjective) approach to form their expectations are more likely to have biased expectations about the likely success of their entrepreneurial efforts. Second, the hypothesis of reference group neglect by Camerer and Lovallo (1999) suggests the possibility that entrepreneurs are optimistic because they pay insufficient attention to potential post-entry competition. We acknowledge that our COMP_ANAL measures are not fully comparable for Finland and the US, but the measures, as constructed, share the idea that some sort of pre-entry analysis and planning has been done.[28]

[Insert Table 4 about here]

Five findings stand out from the first column of Table 4:

First, (absolute) optimism is increasing in dif_OFAIL_IFAIL. The more optimistic an entrepreneur is about the survival of his own venture relative to that of the other ventures, the higher his optimism. The coefficients of dif_OFAIL_IFAIL are significant at the 1% significance level and they imply that as interpersonal optimism (more accurately, one's own placement relative to the average) increases by one percentage point, optimism increases by 0.5 percentage points in the Finnish sample and 0.8 percentage points in the US sample.

Second, OTHERSFAIL obtain a negative and significant coefficient. This means that the higher the perceived exit rate of similar firms, the lower the absolute optimism. Interestingly, D_OTHERSFAIL obtains a positive and significant coefficient. We can interpret this finding to mean that those who are not able or willing to provide a base rate estimate hold more optimistic views.

Third, industry experience reduces optimism. This result squares nicely with the findings of Landier and Thesman (2009) and confirms that experience

[28] By neglecting his reference group, an entrepreneur who chooses to enter ends up in competing with entrepreneurs who all think that they are relatively skilled. Consistent with this, Coelho et al. (2004) report that subjects in their experiment have exaggerated beliefs of their own ability and that they systematically condition their behaviour on those beliefs.

reduces biases also in this behavioral domain.

Fourth, like Arabsheibani, de Meza, Maloney and Pearson (2000), we find that the highly educated are less optimistic both in Finland and in the US. This sharpens Cassar's (2010) findings, but is in contrast to the finding of Landier and Thesman (2009). As we see it, this casts some doubt on the view that the highly educated have better outside options and therefore become entrepreneurs only when they are exceptionally optimistic.

Finally, systematic pre-entry planning (i.e., preparation of projected financial statements) is associated with less optimistic survival views (US), whereas pre-entry analysis of competitive situation is associated with more optimistic views (Finland). These are a bit surprising findings and it is not clear how they can be reconciled with the prior evidence: On the one hand, our US result sharpens Cassar's (2010) findings, as he does *not* find a statistically significant relation when the dependent variable is the difference between a nascent entrepreneur's subjective success probability of his firm becoming operational and a dummy variable for the firm's subsequent status. On the other hand, Cassar does find that the use of plans and financial projections by a nascent entrepreneur is positively correlated with optimistic venture sale forecasts, conditional on entry taking place. Our findings for the US are not consistent with this, whereas the Finnish results are: In particular, the planning fallacy argument of Cassar is line with the Finnish findings, whereas the reference group neglect view of Camerer and Lovallo (1999) is not: If the latter was an important source of positive bias in survival expectations, we would not expect to observe a positive relation in the Finnish data. In sum, we cannot be conclusive on this. Whether entrepreneurial optimism around market entry reflects some sort of planning fallacy or reference group neglect remains to be confirmed.

It also is worth pointing out from Table 4 that the results are robust across the columns. In particular, dropping the entrepreneurs for whom OTHERSFAIL is missing does not change the results.

While the regression results of Table 4 cannot be interpreted to provide conclusive evidence on the (causal) causes of entrepreneurial optimism, they are in our view a useful first step towards understanding its sources. Optimism around market entry appears to be related to the better-than-average phenomenon (i.e., overplacement; see Moore and Healy 2008), estimation and understanding of base rates, experience, as well as education.[29]

2.4 Robustness tests

So far, we have used a measure of absolute entrepreneurial optimism that replaces unobserved $\pi_i(p_i)$ in (1) by the corresponding predicted probability. The argument for using $\widehat{e}_i$ as a consistent estimator of e_i is, of course, conditional on the model for $\pi_i(p_i)$ being correctly specified. While the results we have presented are robust to using different parametric models (Logit, Probit, Cloglog), we present two additional robustness checks that explore whether our main findings depend on how we model $\pi_i(p_i)$.

First, we estimate a semiparametric model for $\pi_i(p_i)$, which calls for less stringent assumptions than the parametric models. In particular, we estimate a univariate binary-choice model using the semi-nonparametric estimator of Gallant and Nychka (1987) and a Hermite polynomial expansion to approximate

[29] An obvious further question is, how is entrepreneurs' perceived or true ability related to their degree of optimism? A positive bias in the perceived ability might well lead to the overestimation of one's chance of success. Unfortunately, our data do not allow us to study this relation in greater detail. We can, however, shed some light on it by using a subset of the Finnish data set, the 2003 survey, as it includes a question that elicited entrepreneurs' subjective views of how good or able entrepreneurs they consider themselves to be (on Likert scale 1 to 10). The pairwise correlation between the self-assessed entrepreneurial ability and our measure of optimism ($\widehat{e}_i$) is 0.31 and statistically significant at 1% level. The correlation is nearly identical (0.32, p-value < 0.01) if we first regress the ability measure on work experience and indicators of education and use the residuals instead of the raw measure. While exploratory, these univariate findings suggest that entrepreneurial optimism is related to entrepreneurs' self-perceptions and particularly to how able they think they are.

the unknown density of the errors (Gabler, Laisney, and Lechner 1993). Using these models, we compute new semi-parametric measures of optimism, $\widehat{e}_i^{snp}$, and repeat our quintile and regression analyses.

The semiparametric results are reported in Table 5. As can be seen from table, the results echo our earlier findings. For example, the medians are nearly identical to the estimates from the parametric models. The table also shows that in both countries, entrepreneurs at the ninth decile of the optimism distribution are 2-3 times more optimistic than those at the median. While these numbers are a bit higher than the corresponding ratios from the parametric models, they are of the same order of magnitude nevertheless.

[Insert Table 5 about here]

Second, following Cassar (2010), we use $\widetilde{e}_i = Y_i - p_i$ as an alternative measure of optimism (see also Das and van Soest 1997, 1999). We can use $\widetilde{e}_i$ as the dependent variable in the regressions, but due to it being a mixture of a discrete and continuous variable, this alternative measure clearly does not allow for a meaningful MTP or quantile analysis. Nor does it allow us to calculate a measure for the 90/50 decile-ratio, which is an indicator of how much cross-sectional variation there is in the degree of optimism. The results of the regressions are reported in Table 6. Again, they echo our earlier findings. For example, we find that absolute optimism is increasing in dif_OFAIL_IFAIL, that OTHERSFAIL obtain a negative and significant coefficient, and that experience reduces absolute optimism.

[Insert Table 6 about here]

For our final robustness check we repeat the regression analyses reported in Table 4 and 6, but with an extended control vector. The additional controls are those used to model the exit of ventures (see Appendix 2 for details of these models) but not included in the estimated models of Table 4 and 6. The new controls include, for example, dummies for gender and marital status, as well as the respondent's age and its square. The results of these regressions (not reported) echo our earlier findings, as neither the point estimates nor their statistical significance change much.

3 Optimism and entry

Given that new entrepreneurs hold positively biased expectations of their survival probability conditional on having made an entry, it seems that there may be too much entry. However, the question of how prevalent optimism is around market entry remains. We therefore use another Finnish data set to complement our main analysis. These data allow us to observe individuals' labour market transitions. It also allows us to construct a measure of optimism based on economic expectations of individuals and the subsequent realizations. This measure is similar to that used by Arabsheibani, de Meza, Maloney and Pearson (2000) and we can use it to explore whether optimism, thus measured, covaries with entry into entrepreneurship.

3.1 Data and variable definitions

3.1.1 Data source and sample

Our data source is Statistics Finland's Income Distribution Statistics (IDS). It is a nationally representative data set, covering private Finnish households and their members. The data set is annual and has a short panel aspect, as each household shows up in the data for two consecutive years. The data for IDS are collected mostly from administrative registers, such as census data, tax registers, and social and pension registers. A part of the data comes from an interview-based database of Statistics Finland, called Income and Living Conditions Survey.[30]

The original data cover years from 1994 to 2008. The initial annual IDS sample consists of about 10 000 households per year, but due to the timing of measurement of the expectations variables (see below) and some missing data, our basic estimating sample consists of 32278 household-year observations and refers to 1995-2008.

3.1.2 Measuring optimism and entrepreneurial entry

We derive a measure for optimism using data on respondent i's financial expectations for year t and the subsequent realization. Expectation, $E_{it|t-1}$, is based on question "*How do you think that the financial situation of your household will develop over the next 12 months (or during year t)*?". The following response categories are allowed: "1= clearly better", "2 = somewhat better", "3 = stays about the same", "4 = somewhat worse", and "5 = clearly worse". The question refers to year t, but was asked in the survey that primarily concerns year $t-1$. Actual outcome, A_{it}, is derived from the re-interview in which i participates. It is based on the question "*How do you think that the financial situation of your household developed in year t*?". It allows for the same response categories as the expectation question.

We stress two aspects of these questions: First, they are nearly identical to those used by Souleles (2004) to examine the financial condition of U.S. households and the households' expectations. Like Souleles, we use the match between the two questions to analyse optimism (forecast errors) at the level of individual respondents. Second, it is important to note that while each respondent shows up in the data for two consecutive periods, we can match $E_{it|t-1}$ with A_{it} only

[30] Sampling for the IDS is based on a rotating panel. Each year, about half are new households; the rest have been interviewed (and included in the IDS) once before; see Hyytinen and Putkuri (2012) for additional details.

for the latter period.

We consider two measures for optimism: The first, denoted OP_{it}, is defined as follows: $OP_{it} = 1$ if $A_{it} - E_{it|t-1} < 0$, $OP_{it} = 2$ if $A_{it} - E_{it|t-1} = 0$, and $OP_{it} = 3$ if $A_{it} - E_{it|t-1} > 0$. The logic of this measure is that if the difference between $A_{it} - E_{it|t-1}$ is positive (negative), it is indicative of optimism (pessimism). If it is zero, the expectation matches with the outcome and we can think of the individual as being a "realist" when forming her expecations.

We can also define a second, somewhat richer measure of optimism. We denote it OPX_{it} and calculate it as follows: $OPX_{it} = 1$ if $A_{it} - E_{it|t-1} < -1$, $OPX_{it} = 2$ if $A_{it} - E_{it|t-1} = -1$, $OPX_{it} = 3$ if $A_{it} - E_{it|t-1} = 0$, $OPX_{it} = 4$ if $A_{it} - E_{it|t-1} = 1$ and $OPX_{it} = 5$ if $A_{it} - E_{it|t-1} > 1$. This measure allows for two levels of optimism (pessimism), as it provides separate categories for moderate and extreme optimists (pessimists).

It is very important to note that the two measures, OP_{it} and OPX_{it}, are *not* intended to mirror what the surveyed individuals think about their future employment status. Instead, we use them, like Arabsheibani, de Meza, Maloney and Pearson (2000) do, as general indicators of how prone a person is to make optimistic (or pessimistic) forecast errors.

Table 7 describes the distribution of respondents in the different optimism categories in our baseline estimating sample. It shows that while two fifths of the respondents have made a forecast error of their financial condition, the respondents are neither optimistic nor pessimistic on average.

[Insert Table 7 about here]

Our measure for entrepreneurial entry is based on the respondent's socioeconomic status at the two consecutive survey rounds. The respondent's socioeconomic status can be used to determine whether she is an entrepreneur (i.e., an employer, a self-employed or other kind of entrepreneur), or if she is working for someone else. To capture transitions into entrepreneurship, we let $NEWENT_{it} = 1$ if the respondent is working for someone else at $t-1$ but is an entrepreneur at t and $NEWENT_{it} = 0$ if the respondent is working for someone else at $t-1$ and at t.[31] The sample mean of $NEWENT_{it}$ is 0.027.

3.2 Main results

Table 8 reports the results of simple linear probability models (LPM) in which the dependent variable is $NEWENT_{it}$. We report the results from LPM models as they allow for an easy interpretation and are in some cases more robust to possible model misspecification; however, we note already here that the results are robust to using other standard binary models, such as Probit (see below).

In the first two columns on the left of Table 8, the explanatory variables of key interest are binary variables derived from OP_{it}. The specification in the leftmost column differs from that of its neighbour because the latter also includes a long vector of control variables (gender, age, lagged income, and dummies for family composition, region of residence, level of education, marital status and sample years). In the two rightmost columns of the table, the explanatory variables of key interest are binary variables derived from OPX_{it}, but otherwise the specifications are similar to those on the left.

The table shows that individuals who make optimistic forecast errors concerning year t are more likely to transit into entrepreneurship during that year than those holding realistic beliefs (which is the omitted category). Moreover, as columns (3) and (4) show, the finding is most pronounced for those holding most optimistic views. This is what we would expect if extreme optimism is correlated with non-prudent entry choices. Interestingly, there also is some evidence that pessimists are more likely to transit into entrepreneurship than those holding realistic beliefs. This evidence is, however, weaker than that for

[31] Those who are entrepreneurs both at $t-1$ and t are excluded from the estimating sample.

optimism.

[Insert Table 8 about here]

The results in Table 8 are robust to not allowing for a separate dummy for the pessimists (i.e., to including them to the comparison group). This specification is of particular interest, as it is very close (but not identical) to a standard difference-in-differences estimator, the treatment being optimism. In this specification, the estimated effect is 0.008 and significant at better than 1% level. Adding the long control vector to the specification does not change this finding.

The results are also robust to using standard binary models, such as the Probit model, instead of the LPM. For example, in the specification corresponding to that of column (2) of Table 8, the Probit marginal effect of optimists is 0.009 and highly significant. Finally, the results do not change if we use an alternative definition for the dependent variable in which the respondent can be working for someone else, be unemployed, be out of the labor force, or have an unknown socioeconomic status prior to transiting into entrepreneurship.

3.3 Discussion

The foregoing analysis establishes that individuals who are optimistic are more likely to transit into entrepreneurship. How about exits from entrepreneurship? Are they related to optimism? As far we are aware, no earlier study has considered this question explicitly.

For this analysis we focus on those individuals in our IDS sample who are entrepreneurs at $t-1$ and then check whether they continue in that state at t or whether they exit entrepreneurship. It turns out that if we allow individuals to exit to any possible labour market state (e.g., to start working for someone else, to become unemployed, or to exit the labor force) after exiting entrepreneurship, optimism is *inversely* related to the likelihood of that an individual leaves entrepreneurship for an alternative occupation. This result suggests that those who hold optimistic views stay longer in entrepreneurship. If we restrict the sample so that individuals can only re-enter the labour market to work for someone else, the effect of optimism is negative and significant without control variables. However, it is negative but statistically insignificant once the control vector, similar to that used in Columns (2) and (4) of Table 8, is added to the model. Even in this specification, the coefficient of the most optimistic individuals (i.e., those with $OPX_{it} = 5$) is nearly significant (p-value = 0.11).

In sum, the findings of our complementary analysis are consistent with the view that there is optimism around market entry and that it has behavioural consequences. They also square with the results of Puri and Robinson (2007), who show that the most optimistic individuals behave less prudently. It has been argued that despite greater risks, the monetary returns to entrepreneurial efforts and investments are negligible and that unrealistic entrepreneurial optimism may be important for entry decisions (see, Hamilton 2000, Moskowitz and Vissing-Jorgensen 2002, and Astebro 2003). The above findings provide a confirming angle to this debate.

4 Conclusions

Most economic models of occupational choice and market entry regard (subjective) expectations to be instrumental both for the entry decision and post-entry expansion. We have complemented the earlier literature, particularly the analyses of Landier and Thesman (2009) and Cassar (2010), by using US and Finnish data on nascent and new entrepreneurs and by developing a new econometric approach that links the probabilistic expectations of the entrepreneurs to the subsequent survival of their ventures.

Our analysis confirms some of the results presented earlier in the literature and generates new insights. First, we find - consistent with Landier and Thesman (2009) and Cassar (2010) - that new entrepreneurs hold substantially optimistic survival beliefs both in the US and in Finland. Second, we show that despite the underlying data sets being somewhat different, entrepreneurial optimism displays many similar empirical patterns in the two countries. For example, we find that both in the US and Finland, there is a great deal of variation in how prudent the survival beliefs are. It seems, in particular, that a number of new entrepreneurs are excessively optimistic and that a rather small subset holds unbiased beliefs. Third, unlike the preceding studies, we are able to document that both in Finland and in the US, entrepreneurial optimism is positively correlated with the relative (better-than-average) optimism. Fourth, we find that having more industry experience is associated with less optimistic expectations both in the Finnish and US data. This finding is consistent with the conventional wisdom that experience reduces nonstandard behavior in markets and squares nicely with the findings of Landier and Thesman (2009). The role of education is, however, less clear, as we find that having more education is associated with less optimistic expectations. This sharpens Cassar's (2010) findings, but is in contrast to the results of Landier and Thesman. Finally, we use a measure of optimism similar to that developed in Arabsheibani, de Meza, Maloney and Pearson (2000) to document that those who make more optimistic forecast errors are more likely to become entrepreneurs than their non-optimistic counterparts. Such entrepreneurs are also less likely to exit, though our evidence for this is statistically weaker.

Taken together, the findings of this paper suggest that some, but not all, entrepreneurs may suffer from harmful optimism around the time of market entry. If this is true, it is unclear whether - and if so how - preferences can be inferred from behaviour in the particular domain of entrepreneurship. Deeper understanding of both entrepreneurs' entry beliefs as well as of the returns to entrepreneurship would particularly benefit policy-makers, who appear to endorse the benefits of entrepreneurship unconditionally. Current industrial and entry policies are rather generic and do not seem to recognize the possibility that at least some new entrepreneurs' beliefs may be distorted in a systematic way.

References

[1] Arabsheibani, G., de Meza, D., Maloney, J. and Pearson, B. (2000). And a vision appeared into them of a great profit: Evidence of self-deception among the self-employed, *Economic Letters*, Vol. 67, pp. 35-41.

[2] Astebro, T. (2003). The return to independent invention: Evidence of risk seeking, extreme optimism or skewness-loving, *Economic Journal*, Vol. 113, pp. 226-239.

[3] Basset, W. F. and Lumsdaine R. L. (2001). Probability limits: Are subjective assessments adequately accurate?, *Journal of Human Resources*, Vol. 36, No. 2, pp. 327-363.

[4] Benjamini, Y., Krieger, A. and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate, *Biometrika*, Vol. 93, pp. 491-507.

[5] Benjamini, Y and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society*, Series B. 57, pp. 289-300.

[6] Bernardo, A. and Welch, I. (2001). On the evolution of overconfidence and entrepreneurs, *Journal of Economics and Management Strategy*, Vol. 10, No. 3, pp. 301-330.

[7] Bolger, F., Pulford, B. and Colman, A. (2008). Market entry decisions: Effects of absolute and relative confidence, *Experimental Psychology*, Vol. 55, No. 2, pp. 113-120.

[8] Brocas, I. and Carillo, J. (2004) Entrepreneurial boldness and excessive investment, *Journal of Economics and Management Strategy*, Vol. 13, No. 2, pp. 321-350.

[9] Brunnermeier, M. and Parker, J. (2005). Optimal expectations, *American Economic Review*, Vol. 95, No. 2, pp. 1092-1118.

[10] Camerer, C. and Lovallo, D. (1999). Overconfidence and excess entry: An experemential approach, *American Economic Review*, Vol. 89, No. 1, pp. 306-318.

[11] Cassar, G. (2010). Are individuals entering self-employment overly-optimistic? An empirical test of plans and projections on nascent entrepreneur expectations, *Strategic Management Journal*, Vol. 31, Issue 8, pp. 822 - 840.

[12] Clark, J. and Friesen, L. (2009). Overconfidence in forecasts of own performance: An experimental study, *Economic Journal*, Vol. 119, pp. 229-251.

[13] Coelho, M., de Meza, D. and Reyneirs, D. (2004). Irrational exuberance, entrepreneurial finance and public policy, *International Tax and Policy Finance*, Vol. 11, pp. 391-417.

[14] Cooper, A., Woo, C. and Dunkelberg, W. (1988). Entrepreneurs' perceived chances for success, *Journal of Business Venturing*, Vol. 3, pp. 97-108.

[15] Coval, J. D. and Thakor, A. V. (2005). Financial intermediation as a beliefs-bridge between optimists and pessimists, *Journal of Financial Economics*, Vol. 75, pp. 535-569.

[16] Das, M. and van Soest, A. (1997). Expected and realized income changes: Evidence from the Dutch socio-economic panel, *Journal of Economic Behaviour and Organization*, Vol. 32, pp. 137-154.

[17] Das, M. and van Soest, A. (1999). A panel data model for subjective information on household income growth, *Journal of Economic Behaviour and Organization*, Vol. 40, pp. 409-426.

[18] DellaVigna, S. (2009). Psychology and economics: Evidence from the field, *Journal of Economic Literature*, Vol. 47, pp. 315-372.

[19] Dushnitsky, G. (2010). Entrepreneurial optimism in the market for technology inventions, *Organization Science*, Vol. 21, Issue 1, pp. 150-167.

[20] Forbes, D. P. (2005). Are some entrepreneurs more overconfident than others?, *Journal of Business Venturing*, Vol. 20, pp. 623-640.

[21] Fraser, S. and Greene, F. J. (2006). The effects of experience on entrepreneurial optimism and uncertainty, *Economica*, Vol. 73, pp. 169-192.

[22] Gabler, S., Laisney, F., and Lechner, M. (1993). Seminonparametric estimation of binary-choice models with an application to labor-force participation, *Journal of Business and Economic Statistics,* Vol. 11, pp. 61-80.

[23] Gallant, A. R., and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation, *Econometrica,* Vol. 55, pp. 363-390.

[24] Grieco, D. and Hogart, R. M. (2009). Overconfidence in absolute and relative performance: The regression hypothesis and Bayesian updating, *Journal of Economic Psychology*, Vol. 30, pp. 756-771.

[25] Hamilton, B. (2000). Does entrepreneurship pay? An empirical analysis of the returns to self-Employment, *Journal of Political Economy*, Vol. 108, No. 3, pp. 604-631.

[26] Hayward, M. L., Shepherd, D. and Griffin (2006). A hubris theory of entrepreneurship. *Management Science*,Vol. 52, No. 2, pp. 160-172.

[27] Hyytinen A. (2003). Loan market equilibrium with difference of opinion and imperfect competition, *Economics Letters,* Vol. 78, pp. 125–129.

[28] Hyytinen, A. and Putkuri, H. (2012). Household optimism and borrowing, work-in-progress, Bank of Finland.

[29] Koellinger, P., Minniti, M. and Schade, C. (2007). "I think I can, I think I can": Overconfidence and entrepreneurial behaviour, *Journal of Economic Psychology*, Vol. 28, pp. 502-527.

[30] Koenker, R.W. (2005): *Quantile Regression*, Econometric Society Monograph Series. New York : Cambridge University Press.

[31] Landier, A. and Thesmar, D. (2009). Financial contracting with optimistic entrepreneurs: Theory and evidence, *Review of Financial Studies,* Vol. 22, No. 1, pp. 117-150.

[32] List, J. (2003). Does market experience eliminate market anomalies? *Quarterly Journal of Economics*, Vol. 118 No. 1, pp. 41-71.

[33] Manove, M, and Padilla, A. J. (1999). Banking (conservatively) with optimists, *Rand Journal of Economics*, Vol. 30, pp. 324-350.

[34] Manski, C. F. (2004). Measuring expectations, *Econometrica*, Vol. 72, No. 5, pp. 1329-1376.

[35] Manski, C. F. and Molinari, F. (2010). Rounding probabilistic expectations in surveys, *Journal of Business and Economic Statistics*, Vol. 28, No 2, pp. 219-231.

[36] de Meza, D. and Southey, C. (1996). The borrower's curse: Optimism, finance and entrepreneurship, *Economic Journal*, Vol. 106, pp. 375-386.

[37] Moore, D. A. and Healy P. J. (2008). The trouble with overconfidence, *Psychological Review*, Vol. 115, No. 2, pp. 502-517.

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

[39] Pajarinen, M., Rouvinen P. and Ylä-Anttila P. (2006). Growth intentions of new entrepreneurs [Uusyrittäjien kasvuhakuisuus], The Research Institute of the Finnish Economy, Discussion Papers 1052.

[40] Puri, M. and Robinson, D. (2007). Optimism and economic choice, *Journal of Financial Economics*, Vol. 86, pp. 71-99.

[41] Puri, M. and Robinson, D. (2009). The economic psychology of entrepreneurship and family business, Duke University.

[42] Rabin, M. (1998). Psychology and economics, *Journal of Economic Literature*, Vol. 36, pp. 11-46.

[43] Rouvinen, P. and Yla-Anttila, P. (2004). Entrepreneurship grows fastest and recruits most in services [Palveluyrittäjyys kasvaa nopeimmin ja työllistää eniten], The Research Institute of the Finnish Economy, Discussion Papers 913.

[44] Santo-Pinto, L. and Sobel, J. (2005). A model of positive self image in subjective assesments, *American Economic Review*, Vol. 95 No. 5., pp. 1386-1402.

[45] Souleles, N. S. (2002). Expectations, heterogenous forecast errors and consumption: micro evidence from the Michigan Consumer Sentiment Surveys, *Journal of Money, Credit and Banking*, Vol. 36, No. 1, pp. 39-72.

[46] Ucbasaran, D., Westhead, P., Wright, M. and Flores, M. (2010). The nature of entrepreneurial experience, business failure, and comparative optimism, *Journal of Business Venturing*, Vol. 25, pp. 541-555.

[47] van den Steen, E. (2004). Rational overoptimism (and other biases), *American Economic Review*, Vol 94 No. 4, pp. 1141-1151.

[48] Viscusi, W. K. (1990). Do smokers underestimate risks? *Journal of Political Economy*, Vol. 98, No. 6, pp. 1253-1269.

Figures and tables

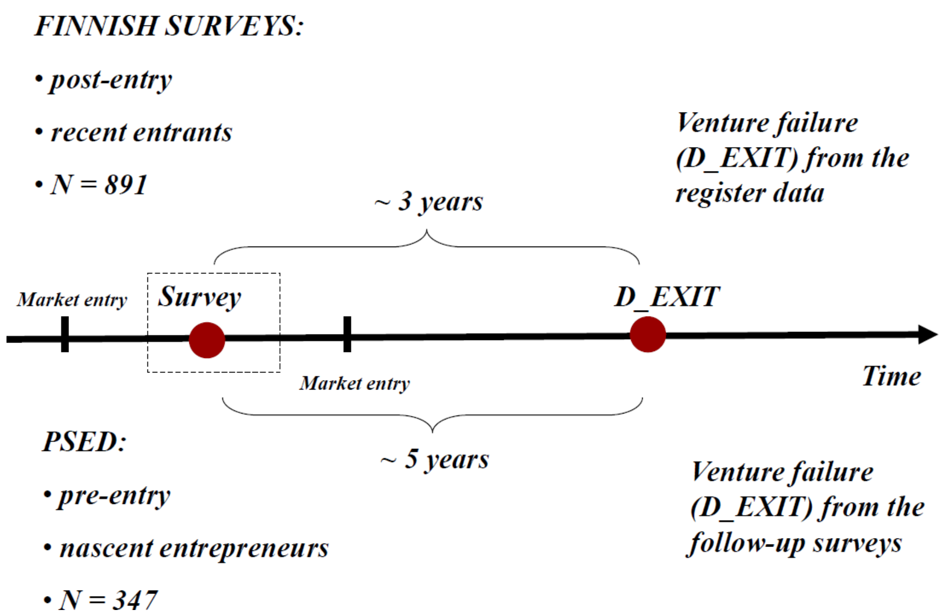

Figure 1: The summary of the data and the timing of measurement.

Table 1

Panel A: The distribution of IFAIL

IFAIL category	FIN Freq.	FIN Percent	FIN Cum.	US Freq.	US Percent	US Cum.
0	307	34.46	34.46	163	46.97	46.97
1-10	317	35.58	70.03	42	12.10	59.08
11-20	115	12.91	82.94	29	8.36	67.44
21-40	74	8.31	91.25	54	15.56	83.00
41-100	78	8.75	100.00	59	17.00	100.00
Total	891			347		

Panel B: Mean of IFAIL and D_EXIT and their correlation

	FIN	p-value	US	p-value
mean of IFAIL	11.86 %		17.14 %	
mean of D_EXIT	34.34 %		65.42 %	
corr(IFAIL, D_EXIT)	0.21	(<0.01)	0.08	(0.15)

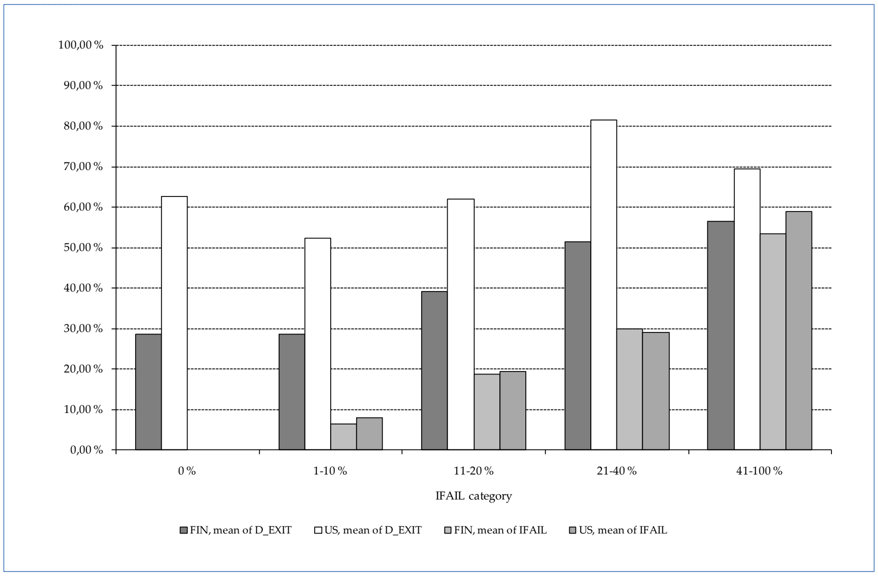

Figure 2: The means of D_EXIT and IFAIL conditional on IFAIL category

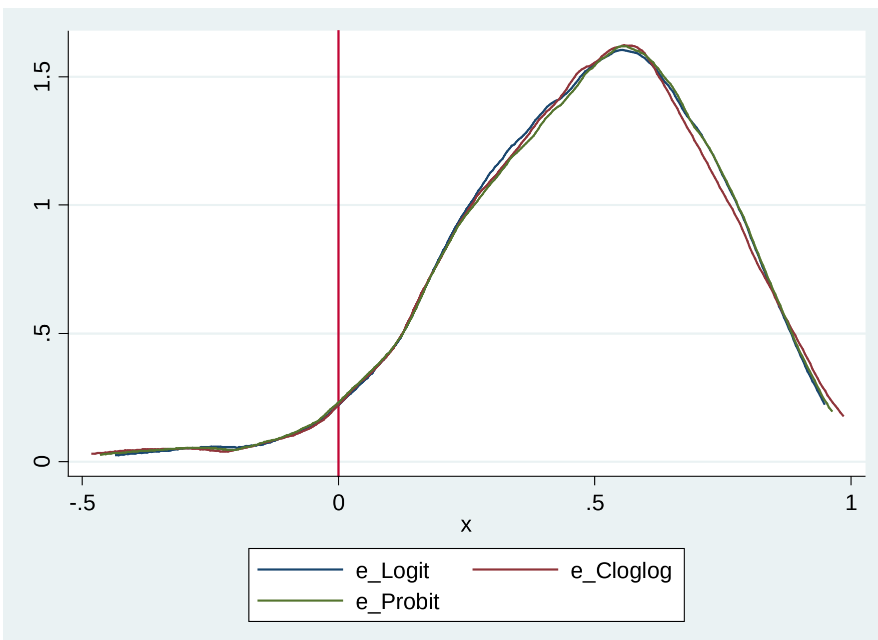

Figure 3: The estimated d.f. of the difference between the objective and subjective probabilities that the venture of an entrepreneur is not in business after five years. (US)

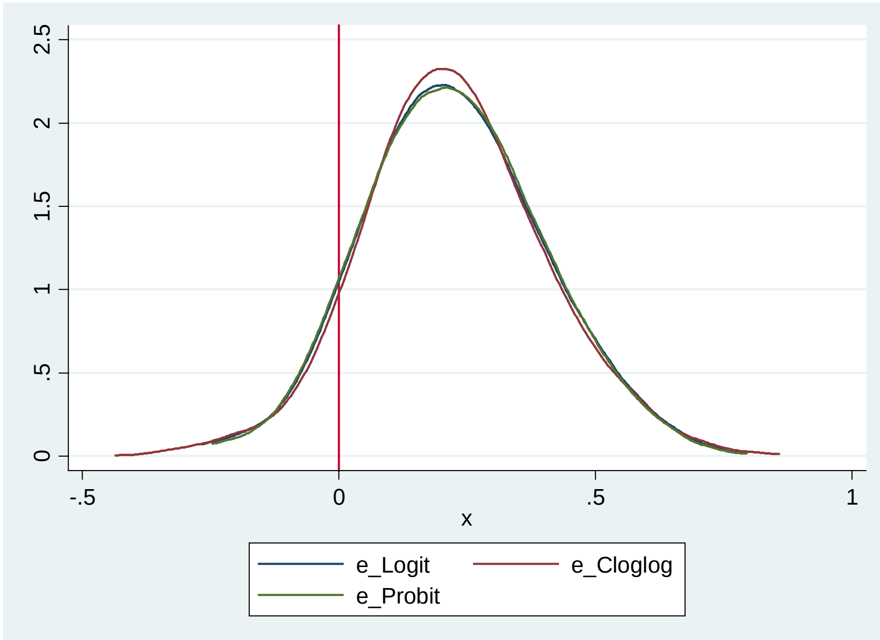

Figure 4: The estimated d.f. of the difference between the objective and subjective probabilities that the venture of an entrepreneur is not in business after three years. (FIN)

Table 2

Quantile regressions

	US		FIN	
	Coef.	S.E.(Robust)	Coef.	S.E.(Robust)
q10	0.179	0.029 ***	0.025	0.007 ***
q20	0.286	0.015 ***	0.083	0.008 ***
q30	0.366	0.022 ***	0.132	0.008 ***
q40	0.441	0.020 ***	0.180	0.006 ***
q50	0.508	0.015 ***	0.214	0.006 ***
q60	0.562	0.018 ***	0.260	0.006 ***
q70	0.624	0.017 ***	0.302	0.008 ***
q80	0.693	0.016 ***	0.364	0.010 ***
q90	0.785	0.017 ***	0.453	0.010 ***
Observations	347		891	

Notes: Dependent variable is the difference between the objective (predicted probability, Logit model) and subjective probabilities that the venture is not in business five (US) or three (FIN) years after entry. The quantile regression model includes only a constant; ***, **, and * denote statistical significance at 1%, 5% and 10% levels, respectively.

Table 3

Classification based on multiple testing procedure (MTP)

US

Group	Means of prediction errors in percentage points	N	Proportion	Average prediction error in percentage points
Pessimist	-0.37	2	0.01	
Realist	0.11	48	0.14	
Optimist	0.55	297	0.86	
Total		347	1.00	0.48

FIN

Group	Means of prediction errors in percentage points	N	Proportion	Average prediction error in percentage points
Pessimist	-0.21	12	0.01	
Realist	0.07	290	0.33	
Optimist	0.31	589	0.66	
Total		891	1.00	0.22

Notes: MTP has three steps: i) test H_0: e=0 for each entrepreneur, ii) compute and order the associated p-values in an increasing order and iii) use a sequence of critical values to identify entrepreneurs for whom the null hypothesis is rejected. The critical values are adjusted so that they control for FDR at 5 %. An entrepreneur is then inferred to hold optimistic (pessimistic) beliefs if the prediction error is positive (negative) and if he belongs to the subsample of entrepreneurs for whom the null hypothesis is rejected.

Table 4

Regression (CEF) results

US	(1)			(2)			(3)			(4)		
	Coef.	S.E.(Robust)		Coef.	S.E.(Robust)		Coef.	S.E.(Robust)		Coef.	S.E.(Robust)	
dif_OFAIL_IFAIL	0.811	0.035	***	0.394	0.049	***	0.800	0.035	***	0.810	0.035	***
OTHERSFAIL	-0.769	0.046	***				-0.759	0.047	***	-0.763	0.046	***
D_OTHERSFAIL	0.104	0.021	***									
EXPERIENCE	-0.011	0.001	***				-0.013	0.001	***	-0.013	0.001	***
D_EXPERIENCE	-0.041	0.016	**				-0.021	0.016		-0.025	0.016	
COMP_ANAL	-0.089	0.026	***							-0.092	0.028	***
ACADEMIC	-0.057	0.027	**	-0.059	0.045		-0.031	0.027		-0.025	0.026	
COM_COLLEGE	-0.108	0.017	***	-0.134	0.028	***	-0.124	0.017	***	-0.120	0.017	***
CONSTANT	0.807	0.033	***	0.333	0.029	***	0.730	0.027	***	0.813	0.035	***
Observations	347			247			247			247		
Adj-R2	0.631			0.316			0.737			0.750		

FIN	(1)			(2)			(3)			(4)		
	Coef.	S.E.(Robust)		Coef.	S.E.(Robust)		Coef.	S.E.(Robust)		Coef.	S.E.(Robust)	
dif_OFAIL_IFAIL	0.478	0.037	***	0.250	0.034	***	0.481	0.037	***	0.482	0.037	***
OTHERSFAIL	-0.344	0.035	***				-0.342	0.037	***	-0.348	0.035	***
D_OTHERSFAIL	0.075	0.012	***									
EXPERIENCE	-0.007	0.001	***				-0.009	0.001	***	-0.008	0.001	***
COMP_ANAL	0.102	0.009	***							0.088	0.010	***
D_2003	0.077	0.012	***	0.006	0.012		0.088	0.013	***	0.091	0.012	***
ACADEMIC	-0.147	0.014	***				-0.131	0.016	***	-0.141	0.015	***
COM_COLLEGE	-0.067	0.011	***				-0.061	0.012	***	-0.071	0.011	***
CONSTANT	0.225	0.011	***	0.179	0.009	***	0.274	0.011	***	0.231	0.011	***
Observations	891			683			683			683		
Adj-R2	0.354			0.088			0.313			0.385		

Notes: Dependent variable is the difference between the objective (predicted probability, Logit model) and subjective probabilities that the venture is not in business five (US) or three (FIN) years after entry. S.E refers to robust standard errors; ***, **, and * denote statistical significance at 1%, 5% and 10% levels, respectively.

Table 5

Quantile regressions and regression (CEF) analysis based on semiparametric prediction errors

a) *Quantile regressions (constant only)*

	US			FIN		
	Coef.	S.E.(Robust)		Coef.	S.E.(Robust)	
q10	0.179	0.029	***	-0.045	0.006	***
q20	0.286	0.015	***	-0.004	0.007	
q30	0.366	0.022	***	0.053	0.013	***
q40	0.441	0.020	***	0.124	0.013	***
q50	0.508	0.015	***	0.193	0.009	***
q60	0.562	0.018	***	0.266	0.014	***
q70	0.624	0.017	***	0.389	0.018	***
q80	0.693	0.016	***	0.511	0.022	***
q90	0.785	0.017	***	0.656	0.013	***
Observations	347			891		

b) Regression results

US	(1)			(2)			(3)			(4)		
	Coef.	S.E.(Robust)		Coef.	S.E.(Robust)		Coef.	S.E.(Robust)		Coef.	S.E.(Robust)	
dif_OFAIL_IFAIL	0.942	0.071	***	0.472	0.071	***	0.923	0.071	***	0.944	0.072	***
OTHERSFAIL	-0.812	0.090	***				-0.795	0.092	***	-0.803	0.090	***
D_OTHERSFAIL	0.087	0.033	***									
EXPERIENCE	-0.018	0.002	***				-0.019	0.002	***	-0.020	0.002	***
D_EXPERIENCE	-0.075	0.027	***				-0.031	0.031		-0.039	0.030	
COMP_ANAL	-0.160	0.043	***							-0.186	0.046	***
ACADEMIC	-0.070	0.046		-0.065	0.075		-0.035	0.057		-0.023	0.054	
COM_COLLEGE	-0.063	0.029	**	-0.077	0.041	*	-0.073	0.032	**	-0.066	0.031	**
CONSTANT	0.960	0.059	***	0.340	0.041	***	0.804	0.051	***	0.971	0.062	***
Observations	347			247			247			247		
Adj-R2	0.444			0.187			0.509			0.535		

FIN	(1)			(2)			(3)			(4)		
	Coef.	S.E.(Robust)		Coef.	S.E.(Robust)		Coef.	S.E.(Robust)		Coef.	S.E.(Robust)	
dif_OFAIL_IFAIL	0.265	0.064	***	0.219	0.053	***	0.265	0.064	***	0.266	0.064	***
OTHERSFAIL	-0.082	0.061					-0.075	0.063		-0.083	0.061	
D_OTHERSFAIL	0.149	0.019	***									
EXPERIENCE	-0.009	0.002	***				-0.010	0.002	***	-0.009	0.002	***
COMP_ANAL	0.145	0.015	***							0.127	0.017	***
D_2003	0.098	0.021	***	0.007	0.020		0.107	0.024	***	0.112	0.023	***
ACADEMIC	-0.237	0.023	***				-0.223	0.028	***	-0.237	0.027	***
COM_COLLEGE	-0.113	0.018	***				-0.108	0.021	***	-0.122	0.021	***
CONSTANT	0.197	0.019	***	0.184	0.014	***	0.267	0.020	***	0.206	0.020	***
Observations	891			683			683			683		
Adj-R2	0.243			0.026			0.145			0.204		

Notes: Dependent variable is the difference between the objective (predicted probability, semiparametric model) and subjective probabilities that the venture is not in business five (US) or three (FIN) years after entry. S.E refers to robust standard errors; ***, **, and * denote statistical significance at 1%, 5% and 10% levels, respectively.

Table 6

Regression (CEF) analysis with D_EXIT-IFAIL as the dependent variable

US	(1)		(2)		(3)		(4)	
	Coef.	S.E.(Robust)	Coef.	S.E.(Robust)	Coef.	S.E.(Robust)	Coef.	S.E.(Robust)
dif_OFAIL_IFAIL	0.896	0.137 ***	0.438	0.118 ***	0.870	0.137 ***	0.882	0.137 ***
OTHERSFAIL	-0.839	0.173 ***			-0.827	0.173 ***	-0.831	0.172 ***
D_OTHERSFAIL	0.104	0.056 *						
EXPERIENCE	-0.012	0.004 ***			-0.011	0.005 **	-0.012	0.005 **
D_EXPERIENCE	-0.042	0.051			-0.032	0.061	-0.037	0.061
COMP_ANAL	-0.092	0.078					-0.105	0.101
ACADEMIC	-0.059	0.088	-0.053	0.108	-0.024	0.111	-0.018	0.111
COM_COLLEGE	-0.104	0.058 *	-0.160	0.071 **	-0.146	0.066 **	-0.142	0.066 **
CONSTANT	0.818	0.108 ***	0.323	0.067 ***	0.743	0.099 ***	0.837	0.124 ***
Observations	347		247		247		247	
R2	0.159		0.088		0.186		0.189	

FIN	(1)		(2)		(3)		(4)	
	Coef.	S.E.(Robust)	Coef.	S.E.(Robust)	Coef.	S.E.(Robust)	Coef.	S.E.(Robust)
dif_OFAIL_IFAIL	0.462	0.127 ***	0.242	0.095 **	0.457	0.125 ***	0.458	0.126 ***
OTHERSFAIL	-0.334	0.125 ***			-0.328	0.124 ***	-0.333	0.125 ***
D_OTHERSFAIL	0.075	0.037 **						
EXPERIENCE	-0.007	0.003 **			-0.008	0.004 **	-0.007	0.004 **
COMP_ANAL	0.102	0.031 ***					0.077	0.035 **
D_2003	0.077	0.039 **	0.002	0.036	0.082	0.044 *	0.085	0.044 *
ACADEMIC	-0.147	0.049 ***			-0.191	0.053 ***	-0.200	0.054 ***
COM_COLLEGE	-0.067	0.036 *			-0.065	0.041	-0.074	0.041 *
CONSTANT	0.225	0.037 ***	0.181	0.025 ***	0.282	0.037 ***	0.244	0.039 ***
Observations	891		683		683		683	
R2	0.045		0.010		0.042		0.049	

Notes: Dependent variable is the difference between the binary outcome and subjective probability for the event that the venture is not in business five (US) or three (FIN) years after entry. S.E refers to robust standard errors; ***, **, and * denote statistical significance at 1%, 5% and 10% levels, respectively.

Table 7

Distribution of OP and OPX optimism measures

Pessimist	Realist	Optimist
OP=1	OP=2	OP=3
0.236	0.538	0.227

Extreme pessimist	Moderate pessimist	Realist	Moderate optimist	Extreme optimist
OPX=1	OPX=2	OPX=3	OPX=4	OPX=5
0.042	0.193	0.538	0.182	0.045

Notes: The sample is derived from Statistics Finland's Income Distribution Statistics and covers year from 1995 to 2008. The number of individual-year observations is 32278.

Table 8

Linear probability models of entrepreneurial entry

	(1)		(2)		(3)		(4)	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
OP = 1	0.005	0.002 **	0.005	0.002 **				
OP = 3	0.009	0.002 ***	0.009	0.003 ***				
OPX = 1					0.004	0.005	0.005	0.005
OPX = 2					0.005	0.002 **	0.005	0.003 **
OPX = 4					0.008	0.003 ***	0.008	0.003 ***
OPX = 5					0.013	0.005 **	0.014	0.006 ***
Controls	No		Yes		No		Yes	
Observations	32728		25786		32728		25786	

Notes: S.E. refers to standard errors that are robust to heteroscedasticity. ***, **, and * denote statsitical significance at 1%, 5% and 10% levels, respectively. The omitted category refers to the group of realists (OP=2 or OPX = 3). The control vector includes gender, age, lagged income, and dummies for family composition, region of residence, level of education, marital status, and years.

Appendix 1 (not for publication): Definitions and descriptive statistics of explanatory variables

Table A1

Descriptive statistics of the US data

Variable	Description	Mean	Median	Std. Dev.	Min	Max	Obs.
D_EXIT	=1 if company has failed in 5 years, 0 otherwise	0.654	1.000	0.476	0.000	1.000	347
IFAIL	Failure prob. of entrepr. own business in the next 5 yrs.	0.171	0.050	0.227	0.000	0.950	347
OTHERSFAIL	Failure rate of similar firms during the next 5 yrs.	0.593	0.600	0.221	0.000	1.000	247
D_OTHERSFAIL	= 1 if no response to OTHERSFAIL probability, 0 otherwise	0.288	0.000	0.454	0.000	1.000	347
ACADEMIC	= 1 if respondent has an academic degree, 0 otherwise	0.092	0.000	0.290	0.000	1.000	347
COM_COLLEGE	= 1 if respondent has a community college level degree, 0 otherwise	0.305	0.000	0.461	0.000	1.000	347
AGE	The age of the respondent in years/10	4.073	4.000	1.083	1.800	7.400	347
FEMALE	= 1 for females and 0 for males	0.542	1.000	0.499	0.000	1.000	347
SINGLE	= 1 if the respondent was single at the time of interview, 0 otherwise	0.202	0.000	0.402	0.000	1.000	347
NCHILD	The number of children in the respondent's household	1.372	1.000	1.508	0.000	9.000	347
EXPERIENCE	Sums respondent's work and entrepr. exper. in the field of the start-up in yrs.	8.485	5.000	9.503	0.000	50.000	165
D_EXPERIENCE	1 if no response to EXPERIENCE, 0 otherwise	0.524	1.000	0.500	0.000	1.000	347
TEAM	= 1 if several founders, 0 otherwise	0.553	1.000	0.498	0.000	1.000	347
COMP_ANAL	=1 if respondent gathered information about competition, =0 otherwise	0.879	1.000	0.327	0.000	1.000	347

Table A2

Descriptive statistics of the Finnish data

Variable	Description	Mean	Median	Std. Dev.	Min	Max	Obs.
D_EXIT	=1 if company has failed in 3 years, 0 otherwise	0.343	0.000	0.475	0.000	1.000	891
IFAIL	Failure prob. of entrepr. own business in the next 3 yrs.	0.119	0.050	0.161	0.000	0.990	891
OTHERSFAIL	Failure rate of similar firms during the next 3 yrs.	0.240	0.200	0.193	0.000	0.900	683
D_OTHERSFAIL	= 1 if no response to OTHERSFAIL probability, 0 otherwise	0.233	0.000	0.423	0.000	1.000	891
ACADEMIC	= 1 if respondent has an academic degree, 0 otherwise	0.126	0.000	0.332	0.000	1.000	891
COM_COLLEGE	= 1 if respondent has a community college level degree, 0 otherwise	0.488	0.000	0.500	0.000	1.000	891
AGE	The age of the respondent in years/10	3.673	3.600	1.027	1.900	6.800	891
FEMALE	= 1 for females and 0 for males	0.365	0.000	0.482	0.000	1.000	891
SINGLE	= 1 if the respondent was single at the time of interview, 0 otherwise	0.202	0.000	0.402	0.000	1.000	891
NCHILD	The number of children in the respondent's household	1.083	1.000	1.287	0.000	8.000	891
WORKED	= 1 if respondent was employed prior to starting the new business, 0 otherwise	0.793	1.000	0.405	0.000	1.000	891
EXPERIENCE	Sums respondent's work and entrepr. exper. in the field of the start-up in yrs.	3.251	1.000	5.950	0.000	40.000	891
TEAM	= 1 if several founders, 0 otherwise	0.891	1.000	0.972	0.000	8.000	891
EMP_SIZE	Number of employees in the start-up	1.749	1.000	2.033	0.000	22.000	891
RETIME	The time (in seconds/10) it took for the respondent to answer to IFAIL	1.171	1.252	1.040	0.012	7.712	891
LOTTERY	Payment in a lottery that has 10% prob. to win 1000e and 90% prob. to win 0e	25.074	5.000	65.421	0.000	500.000	891
RISKLOVE	= 1 if loves taking risks, 0 otherwise	0.506	1.000	0.500	0.000	1.000	891
COMP_ANAL	=1 if respondent gathered information about competition, =0 otherwise	0.511	1.000	0.500	0.000	1.000	891
GROWTH_INTENT	Expected increase in the number of employees during the next 3 yrs.	1.685	0.000	4.083	-2.000	48.000	891
D_2003	=1 if start-up in the 2003 sample, 0 if in the 2005 sample	0.406	0.000	0.491	0.000	1.000	891

Appendix 2 (not for publication): Estimated Logit, Probit and Cloglog models for venture exits

Table A3

Binary regression results (US)

DEPENDENT VARIABLE: D_EXIT	Logit			Probit			Cloglog		
	(1)			(2)			(3)		
	Coef.	S.E.(Robust)		Coef.	S.E.(Robust)		Coef.	S.E.(Robust)	
Constant	4.163	2.917		1.577	1.237		2.149	1.770	
ifail_notzero	-2.049	2.321		-0.371	0.727		-1.381	1.441	
F_inv(p)	0.156	0.154							
ACADEMIC	-0.234	0.439		-0.138	0.261		-0.142	0.261	
COM_COLLEGE	-0.564	0.274	**	-0.357	0.165	**	-0.387	0.170	**
AGE	-0.775	0.828		-0.442	0.466		-0.389	0.435	
AGE2	0.099	0.094		0.057	0.053		0.052	0.051	
FEMALE	-0.257	0.265		-0.145	0.156		-0.132	0.155	
SINGLE	0.618	0.358	*	0.369	0.206	*	0.376	0.195	*
NCHILD	0.153	0.084	*	0.094	0.051	*	0.118	0.052	**
OTHERSFAIL	0.257	0.661		0.170	0.400		0.125	0.418	
D_OTHERSFAIL	0.381	0.277		0.238	0.164		0.254	0.159	
EXPERIENCE	-0.055	0.021	***	-0.034	0.012	***	-0.038	0.014	***
D_EXPERIENCE	-0.154	0.288		-0.086	0.173		-0.063	0.172	
TEAM	0.056	0.283		0.028	0.170		0.030	0.170	
COMP_ANAL	-0.377	0.399		-0.240	0.235		-0.256	0.218	
Industry-dummies	Yes			Yes			Yes		
Observations	347			347			347		
Log pseudolikelihood	-202.032			-202.101			-200.882		

Notes: S.E. refers to robust standard errors; ***, ** and * denote statistical significance at 1%, 5% and 10% levels, respectively.

Table A4

Binary regression results (FIN)

DEPENDENT VARIABLE: D_EXIT	Logit		Probit		Cloglog	
	(1)		(2)		(3)	
	Coef.	S.E.(Robust)	Coef.	S.E.(Robust)	Coef.	S.E.(Robust)
Constant	7.430	1.767 ***	3.316	0.855 ***	5.766	1.504 ***
ifail_notzero	-4.112	1.118 ***	-1.382	0.376 ***	-3.408	1.016 ***
F_inv(p)	0.304	0.077 ***	0.365	0.087 ***	0.254	0.071 ***
ACADEMIC	-0.597	0.284 **	-0.375	0.169 **	-0.415	0.231 *
COM_COLLEGE	-0.353	0.185 *	-0.223	0.111 **	-0.262	0.143 *
AGE	-1.718	0.610 ***	-0.991	0.359 ***	-1.348	0.481 ***
AGE2	0.200	0.076 ***	0.116	0.045 **	0.156	0.061 **
FEMALE	-0.161	0.189	-0.102	0.113	-0.137	0.151
SINGLE	-0.158	0.200	-0.103	0.121	-0.102	0.156
NCHILD	0.091	0.070	0.054	0.042	0.070	0.054
WORKED	-0.286	0.193	-0.180	0.118	-0.213	0.142
EXPERIENCE	-0.018	0.017	-0.011	0.010	-0.013	0.015
TEAM	-0.174	0.147	-0.086	0.085	-0.155	0.120
EMP_SIZE	-0.191	0.140	-0.122	0.077	-0.122	0.122
EMP_SIZE2	-0.005	0.013	-0.001	0.006	-0.008	0.013
RETIME	0.125	0.125	0.076	0.075	0.090	0.093
LOTTERY	-0.001	0.001	-0.001	0.001	-0.001	0.001
RISKLOVE	-0.072	0.160	-0.037	0.095	-0.052	0.125
OTHERSFAIL	0.803	0.503	0.432	0.300	0.669	0.392 *
D_OTHERSFAIL	0.336	0.178 *	0.191	0.107 *	0.304	0.137 **
GROWTH_INTENT	0.027	0.020	0.014	0.012	0.023	0.016
COMP_ANAL	0.476	0.164 ***	0.298	0.097 ***	0.325	0.128 **
D_2003	0.054	0.332	0.053	0.198	0.016	0.259
Industry-dummies	Yes		Yes		Yes	
Observations	891		891		891	
Log pseudolikelihood	-505.264		-504.996		-505.461	

Notes : S.E. refers to robust standard errors; ***, ** and * denote statistical significance at 1%, 5% and 10% levels, respectively.

BANK OF FINLAND RESEARCH
DISCUSSION PAPERS

ISSN 1456-6184, online

1/2012 Maria Teresa Punzi **Housing market and current account imbalances in the international economy.** 2012. 27 p. ISBN 978-952-462-784-9, online.

2/2012 Luisa Lambertini – Caterina Mendicino – Maria Teresa Punzi **Expectations-driven cycles in the housing market.** 2012. 61 p. ISBN 978-952-462-785-6, online.

3/2012 George A. Waters **Quantity rationing of credit.** 2012. 26 p. ISBN 978-952-462-786-3, online.

4/2012 Karlo Kauko **Why is equity capital expensive for opaque banks?** 2012. 26 p. ISBN 978-952-462-787-0, online.

5/2012 Kaushik Mitra – George W. Evans – Seppo Honkapohja **Fiscal policy and learning.** 2012. 32 p. ISBN 978-952-462-788-7, online.

6/2012 Ian W. Marsh – Wolf Wagner **Why is price discovery in credit default swap markets news-specific?** 2012. 41 p. ISBN 978-952-462-789-4, online.

7/2012 Katja Taipalus **Signaling asset price bubbles with time-series methods.** 2012. 47 p. ISBN 978-952-462-790-0, online.

8/2012 Paavo Miettinen **Information acquisition during a Dutch auction.** 2012. 22 p. ISBN 978-952-462-791-7, online.

9/2012 Diego Moreno – Tuomas Takalo **Optimal bank transparency.** 2012. 33 p. ISBN 978-952-462-792-4, online.

10/2012 Alina Barnett – Martin Ellison **Learning by disinflating.** 2012. 20 p. ISBN 978-952-462-795-5, online.

11/2012 Bill Francis – Iftekhar Hasan – Qiang Wu **Do corporate boards affect firm performance? New evidence from the financial crisis?** 2012. 55 p. ISBN 978-952-462-796-2, online.

12/2012 Bill Francis – Iftekhar Hasan – Liang Song **Are firm- and country-specific governance substitutes? Evidence from financial contracts in emerging markets.** 2012. 55 p. ISBN 978-952-462-797-9, online.

13/2012 Chung-Hua Shen – Yu-Li Huang – Iftekhar Hasan **Asymmetric benchmarking in bank credit rating.** 2012. 40 p. ISBN 978-952-462-798-6, online.

14/2012 Bill Francis – Iftekhar Hasan – Qiang Wu **Corporate governance and cost of bank loan.** 2012. 49 p. ISBN 978-952-462-799-3, online.

15/2012 Isabelle Distinguin – Iftekhar Hasan – Amine Tarazi **Predicting rating changes for banks: How accurate are accounting and stock market indicators?** 2012. 34 p. ISBN 978-952-462-800-6, online.

16/2012 Bill Francis – Iftekhar Hasan – Liang Song – Bernard Yeung **What determines bank stock price synchronicity? Global evidence.** 2012. 46 p. ISBN 978-952-462-803-7, online.

17/2012 Bill Francis – Iftekhar Hasan – Kose John – Maya Waisman **Urban agglomeration and CEO compensation.** 2012. 64 p. ISBN 978-952-462-804-4, online.

18/2012 Manthos D. Delis – Iftekhar Hasan – Pantelis Kazakis **Bank regulations and income inequality: Empirical evidence.** 2012. 45 p. ISBN 978-952-462-805-1, online.

19/2012 Iftekhar Hasan – Tania De Renzis – Heiko Schmiedel **Retail payments and economic growth.** 2012. 37 p. ISBN 978-952-462-806-8, online.

20/2012 Ari Hyytinen – Jukka Lahtonen – Mika Pajarinen **Entrepreneurial optimism and survival.** 2012. 49 p. ISBN 978-952-462-807-5, online.

http://www.suomenpankki.fi/en • email: Research@bof.fi
ISBN 978-952-462-807-5, ISSN 1456-6184, online