DIW Berlin

German Institute
for Economic Research

Discussion Papers

465

Philipp Köllinger
Maria Minniti
Christian Schade

Entrepreneurial Overconfidence: Evidence from a C.A.R.T. Approach

Berlin, January 2005

Opinions expressed in this paper are those of the author and do not necessarily reflect views of the Institute.

DIW Berlin

German Institute
for Economic Research

Königin-Luise-Str. 5
14195 Berlin,
Germany

Phone +49-30-897 89-0
Fax +49-30-897 89-200

www.diw.de

ISSN 1433-0210 (Druck) 1619-4535 (elektronisch)

Philipp Köllinger*

Maria Minniti**

Christian Schade***

Entrepreneurial Overconfidence: Evidence from a C.A.R.T. Approach

Berlin, 2.12.2004

* German Institute for Economic Research (DIW Berlin), Königin-Luise-Str. 5, 14195 Berlin, Germany; Email: pkoellinger@diw.de, Tel.: +49-30-89789-618

** Babson College, Economics Division, Luksic Hall #304, Babson Park, MA, 02457 USA; Email: minniti@babson.edu, Tel.: +1-781-239-4296

*** Humboldt-Universität zu Berlin, Institute for Entrepreneurial Studies and Innovation Management, Ziegelstr. 13a, 10117 Berlin, Germany; Email: schade@wiwi.hu-berlin.de, Tel.: +49-30-2093-5904

We thank the GEM Consortium for granting us the use of the data. Many thanks go to Tobias Krebs for providing programming assistance. Maria Minniti gratefully acknowledges financial support from the Haniel Foundation. All errors are our.

Abstract

We use a sample of 18 countries to study what variables have a significant impact on an individual's decision to start a new business and classification and regression trees for an accurate interpretation of the data. Our results support existing literature suggesting the existence of strong country effects. In addition, we find strong evidence that perceptual variables, such as one's belief about her own skills and ability and about the risk involved in the venture, have a crucial impact on new business creation across all countries in our sample. Our findings are consistent with the idea that entrepreneurs evaluate their businesses by taking an "inside view" of their situation, overestimate their likelihood of success, and, as a result, rely significantly on perceptions rather than on objective expectations of success.

Keywords: CARTs, Entrepreneurship, Overconfidence, Self-Employment

JEL Classifications: J4, D0

1 Introduction

Using a sample based on surveys conducted in 18 countries, we study what variables have a significant impact on an individual's decision to start a new business. Our paper uses classification and regression trees (CARTs) for an accurate interpretation of the data. In particular, CARTs are used for the identification of relevant, typical clusters of individuals that exhibit significant differences in their likelihood of being entrepreneurs. CART, a nonparametric regression and classification method originally introduced by Breiman et. al. (1984) has a number of advantages over traditional parametric regression methods because it allows the relaxation of underlying assumptions, reveals interactions of covariates, and uses them to improve the quality of the model. In addition, CART is robust to outliers and, unlike parametric models, invariant to monotone transformations of predictors (Gatnar 2002).

Our results support existing literature suggesting the existence of strong country effects. In addition, we find strong evidence that perceptual variables have a crucial impact on new business creation across all countries in our sample. Indeed, our CART analyses suggest that the subjective perception of having the sufficient skills, knowledge and ability to start a new business, are the main drivers of the decision to start a new venture. Our findings are consistent with the idea that entrepreneurs evaluate their businesses by taking an "inside view" of their situation, overestimate their likelihood of success, and, as a result, rely significantly on their perceptions rather than on objective expectations of success.

The rationale behind such a behavior is that entrepreneurs have a strong tendency to consider their situation as unique. Once they identify a profit opportunity, they isolate their present situation, namely starting a new business, and treat it as an original and unrepeatable event. As a result, they neglect the available statistics of past and future similar situations that could help them in forming more accurate forecasts of their likelihood of success. Kahneman and Lovallo (1993) define a situation in which forecasting individuals focus on the case at hand as the "inside view." In the inside view, the way to think about a problem is to consider all that one knows about it, with special attention to its unique features. In an alternative, Kahneman and Lovallo also define the "outside view," as the one in which forecasting individuals focus on the statistics of a class of cases chosen to be similar, in relevant ways, to the current situation. Under certain conditions, potential entrepreneurs tend to base their decisions on the predictions generated by the inside view. This suggests not only that entrepreneurs form their decisions largely on perceptions, but also that such perceptions may be overoptimistic and nor related to actual measures of risk or abilities.

To our knowledge, very few empirical studies have combined the insight that potential entrepreneurs react to economic incentives with the insight that subjective perceptions also influence their behavior (Busenitz and Barney, 1997; Cooper et al. 1988). Our paper contributes to the elimination of this gap in the literature and, as a result, contributes to the economic theory of entrepreneurial motivation. Our data are exceptionally well suited for the purpose since they record individuals who are in the process of starting a new business at a particular point in time without being the results of ex post evaluations of past decisions.

2 Literature Review

For a long time, scholars working with analytical models have neglected entrepreneurship and simply treated it as part of the residuals that cannot be attributed to any measurable productive input (Baumol 1993, 1983). Recently, however, entrepreneurship has been modeled explicitly as a form of human capital linked to the long run size of the firm (Bates 1990, Iyigun and Owen 1998, Otani 1996). Thus, at the micro-economic level, most of the work related to entrepreneurship has focused on employment status choices and on the alternative motivations that cause some individuals to select entrepreneurship.

In general, the availability of external financing has been shown to be a crucial determinant of the amount of entrepreneurial activity in a community (Evans and Jovanovic (1989), Evans and Leighton (1989), Kihlstrom and Laffont (1979)). The evidence suggests that entrepreneurs face liquidity constraints and that individuals with greater family wealth are more likely to switch from employment to self-employment. Conditions in the labor market have been also identified as an important determinant of employment status choice though the nature of the relationship is still under debate. For example, Bogenhold and Staber (1991) and Evans and Leighton (1989) found evidence of a positive relationship between unemployment and self-employment. On the other hand, Blanchflower and Oswald (1990), and Taylor (1996) found evidence of a negative relationship between the two variables. Most likely, both effects co-exist and their relative dominance is contingent upon other macroeconomic circumstances.

Age and gender have been also shown to play some role. In fact, the probability of starting a new business increases with age up to a threshold point and decreases thereafter (Levesque and Minniti forthcoming) and male are more likely to start a new business than women (Blanchflower 2004, Reynolds et al. 2003). Surprisingly, education has been shown to be negatively related to the probability of being self-employed, except in some rich countries where post graduate training has been shown to have some positive effects (Blanchflower 2004, Reynolds et al. 2003)).

Although some questions about the direction of the causal relationships remain, there is general agreement that the factors mentioned above all have a systematic impact on individuals' choices with respect to entrepreneurship. Nevertheless, evidence from empirical studies suggests that significant differences exist in the levels of new firm creation across countries and over time (Blanchflower, 2004; Reynolds et al., 2003). These differences may be the result of country specific factors that influence employment status choices at the local level or, more likely, of the way in which a set of interdependent factors interact with each other to form a complex web of incentives and information which, ultimately, individuals make their choices. In fact, in addition to economic and demographic factors, it has been shown that individuals' employment status choices differ also because of the environment in which they are formed and because of information and perception asymmetries (Chell and Baines, 2000).

The emphasis on information and perceptions is not new in economic theories of entrepreneurship. Kirzner (1973, 1979) argues that entrepreneurship is "alertness." That is, the ability to perceive unexploited opportunities. Along similar lines, Hayek (1952) argues that attention is always directed to things that we are on the lookout for and that, as a result, we are able to perceive more clearly. This means that entrepreneurial discovery is not a pure bolt from the blue but it is based on an individual's ability to perceive an unexploited opportunity and act upon it. Among other things, perceptions are molded by an individual's network, risk propensity and confidence in one's skills and abilities.

Cooper et al. (1988) surveyed new entrepreneurs and asked what they believed to be their chances of success. In addition, respondents were asked to estimate the average rate of success for businesses similar to their own. Self-perceived chances of success were uncorrelated to "objective" measures of potential success such as having an adequate education or initial capital. More than 80% of the respondents perceived their likelihood of success to exceed 70

% and about 30% of them claim to be sure of succeeding. In contrast, the chance of success that the surveyed entrepreneurs assigned to businesses similar to their own was, in average, only 59%.

In this paper, we use a large cross country sample to study to what extent, if at all, self-perceptions are part of an individuals' decision to start a new business and if such a phenomenon is systematic across countries. In the analysis, we also control for household income, education level, gender, age, and working status of the individual as alternative explanations of entrepreneurial propensity.

3 The Data

Data used for this analysis originates from the 2001 population survey of the Global Entrepreneurship Monitor (GEM). Initiated in 1999 with 10 participating countries, the scope of the project expanded to 29 countries in 2001. A harmonized, representative population survey with at least 2,000 observations was conducted in each of the participating countries between June and July 2001. The main purpose of the survey was to allow the identification of individuals that

(1) claim they are starting a new firm for themselves or their employers,

(2) expect to own all or part of the new firm, and

(3) currently own and manage a firm.

For our purposes, complete data sets were available for 18 countries with more than 22,000 observations. Countries included in our study are Argentina, Canada, Denmark, Finland, Germany, Hungary, India, Israel, Italy, Japan, New Zealand, Poland, Portugal, Russia, Singapore, South Korea, Sweden, and USA.

Entrepreneurial activity is measured using the total entrepreneurial activity (TEA) index. The TEA index is calculated adding individuals engaged in the start-up process and those managing/owning a new firm. All respondents were asked:

1a. You are, alone or with others, currently trying to start a new business, including any type of self-employment (yes, no, don't know, refuse).

1b. You are, alone or with others, trying to start a new business or a new venture with your employer - an effort that is part of your normal work (yes, no, don't know, refuse).

1c. You are, alone or with others, the owner of a company you help manage (yes, no, don't know, refuse).

Respondents who answered "yes" to items 1a or 1b were then asked:

2a. Over the past twelve months have you done anything to help start this new business, such as looking for equipment or a location, organizing a start-up team, working on a business plan, beginning to save money, or any other activity that would help launch a business? (yes, no, don't know, refuse)

2b. Will you personally own all, part, or none of this business? (all, part, none, don't know, refuse)

2c. Has the new business paid any salaries, wages, or payments in kind, including your own, for more than three months? (yes, no, don't know, refused)

Individuals were coded as "nascent entrepreneurs", if they answered "yes" to 2a and 2b, and "no" to 2c.

In order to distinguish between individuals involved in a "start-up" (nascent criteria) and those involved in a "new business", follow-on questions were asked to those who answered "yes" to item 1c. Specifically:

3a. Do you personally own all, part, or none of this business? (all, part, none, don't know, refuse)

3b. What was the first year the owners received wages, profits, or payments in kind? (4 digit year, or no profits yet, don't know, refuse)

Individuals were coded as "new business owners" if they answered yes to 3a and if the business had paid wages for a period between 3 and 42 months. The sum of nascent entrepreneurs and new business owners as a percentage of surveyed individuals yields the TEA index used as dependent variable in our analysis. Table 1 shows the un-weighted TEA index for all 18 countries in our sample.

Table 1 about here

The independent variables used in the analysis are described in Table 2. All items were part of the GEM adult population survey questionnaire and were asked to each respondent, independently from whether she was involved in entrepreneurial activities.

Table 2 about here

Finally, one of the main advantages of the GEM data, in addition to their timeliness, is the simplicity of the wording with which they have been collected. Although survey data may be affected by several shortcomings, the GEM survey, because of its construction has the merit of simplicity and, as a result, the ability to insure consistency and comparability across countries. In addition, since our main goal is to test for the role of self-perceptions, the use of survey data seem particularly appropriate.

4 CARTs

Originally introduced by Breiman et. al. (1984), CART is a combination of non-parametric regression and cluster analysis. The method allows the segregation of individuals into clusters exhibiting significantly different entrepreneurial probabilities and characteristics. Although often used in medical research (Zhang and Bracken, 1995; Zhang and Singer, 1999), its application to economics problems is still novel.

CART is particularly well suited for our purposes because it identifies significant predictors and detects higher order interdependencies between co-variables while avoiding multicollinearity problems. By simultaneously identifying significant predictors and clusters that exhibit significant differences with respect to the dependent variable, CART provides unique insights into typical characteristics of entrepreneurs. The result, a "tree" presented in graphic form, is both parsimonious and easy to interpret. In the first step, the sample is systematically sorted into completely homogeneous subsets until a saturated tree is found. In our case, complete homogeneity means that a node contains either only entrepreneurs or non-entrepreneurs. The process of splitting nodes is continued and the partition made finer and finer as the layer gets deeper and deeper. This is a hierarchical process that reveals interdependencies between covariates. The process is continued until the nodes are completely homogeneous and cannot be split any further. The result is a saturated tree. The saturated tree is usually too large to be useful and the resulting model is subject to severe over-fitting problems. Thus, we use a cost-complexity pruning algorithm suggested by Breiman et al. (1984) to identify the relevant set of nested trees. Once the set is identified, we use a 20-fold cross validation procedure to select the best classifying tree. Finally, we run significance tests for all splits in the final trees and drop those that are not significant at least at the 95% confidence level. The Appendix presents the mathematical derivations of our method.

We calculate the final best classifying tree for each and all 18 countries in our sample. Figure 1 shows the final classifying tree for the U.S. and it is used below to explain how trees should be interpreted.

Figure 1 about here

In each tree node the number of entrepreneurs (bottom) and non-entrepreneurs (top) is given, as well as the ratio of entrepreneurs (percentage figure above the node). The variable names below the nodes are the predictors that provide the best split for that node. Terminal nodes are numbered in a descending order, with terminal node 1 representing the cluster of individuals with the highest probability to become an entrepreneur. The final US tree consists of 6 terminal nodes. CART uses 5 different predictor variables to construct the tree, namely SUSKILL, OPPORT, AGE, KNOWENT, FAMFUTUR. The root node shows that the unweighted total population of entrepreneurs in the US in 2001 is 9.1%. Each terminal node contains a different number of individuals. Some of the nodes are rather small and describe rare but statistically significant sub-groups (like number 6, which contains only 107 individuals or 3.6% of the sample), whereas others are very large (like number 5, which contains 1,302 individuals or 44.1% of the sample).

The impact of each of the predictor variables on the ratio of entrepreneurs can be followed along the tree branches. For example, the fraction of entrepreneurs increases from 9.1% (root node) to 14.7%, if people believe to have the sufficient skills, knowledge and ability to start their own business. It again increases sharply, to 24.7%, if these people in addition also expect to find good opportunities to start a new business in the next six months in the area where they live. These two splits identify the group of U.S. respondents with the highest probability to start a new business, indeed, 24.7%. This cluster is actually quite large and contains 709 individuals or 24% of the sample, with 175 respondents actually involved in starting a new business. On the other side, if people do not believe to have the necessary skills, knowledge, and ability, the ratio of entrepreneurs drops dramatically from 9.1% to 1.9% (terminal node 5).

Interestingly, for this big cluster of individuals (1,302), there is no more significant variable to further differentiate between entrepreneurs and non-entrepreneurs. This emphasizes further the importance of the SUSKILL variable: In the US, the belief of possessing the necessary skill, knowledge and ability to start a new business seems to be a necessary, though not sufficient, condition for actually doing so. The cluster with the lowest probability to become an entrepreneur is number 6. It consists of individuals who believe to have the necessary skills, but do not currently see good business opportunities and are at least 68 years old. The combination of these factors results in a 0% likelihood of starting a new business.

Moving further down this branch of the U.S. tree, if people are younger than 67.5 years, knowing another entrepreneur and the perception of the family future become the most important variables. The cluster with the second highest probability of individuals starting new businesses (terminal node 2) consists of people who do not expect good business opportunities in the near future, but believe to have sufficient skills, are younger than 67.5 years, personally know another entrepreneur, and believe that their family will be financially better off in 12 months from now. Individuals in this rather small group (213 individuals, or 7.2% of the sample) have a propensity of 16% to become entrepreneurs. All other nodes and country trees can be interpreted in an analogous manner.

Figure 2, Figure 3, and Figure 4 show the trees for Singapore, Denmark, and Japan respectively.

Figure 2 about here

Figure 2 shows the CART for Singapore. 5.9% of the individuals in the root node are entrepreneurs. In parallel to the US tree, SUSKILL is the best predictor in the first layer of the model. About one fourth of the sample said that they believe to have the sufficient skills, knowledge, and experience to start a new business. 17.5% in this group are entrepreneurs, while otherwise the ratio drops to only 2%. The best splits in the second layer of the tree are KNOWENT for the cluster that does not have SUSKILL, and GEMWORK for the other cluster. The highest proportion of entrepreneurs is found in the cluster that contains individuals who have SUSKILL and currently hold a part-time or full-time job (20.2%). The lowest proportion (0.9%) is found in the cluster that contains individuals who do not believe to have the sufficient skills and do not know another entrepreneur. Interestingly, knowing another entrepreneur increases the proportion of entrepreneurs in the cluster that does not have SUSKILL from 0.9% to 7.7%.

Figure 3 about here

Figure 3 shows the tree model for Denmark and is very representative of the tree models for other European countries. The SUSKILL variable is again the most powerful differentiator between entrepreneurs and non-entrepreneurs, and the overall level of entrepreneurial activity is close or slightly below the mean of TEA ratios for all countries in the sample. In Denmark, 5.6% of the sample count as entrepreneurs. Having confidence in one's own skills and abilities raises the chance to become an entrepreneur to 12%. However, if individuals are above 63.5 years of age, their probability to become entrepreneurs drops to zero, even if they are confident in their skills. These two variables (SUSKILL and AGE) are the only significant predictors in the CART model for Denmark.

Figure 4 about here

Finally, Figure 4 shows the CART model for Japan, the country with the lowest proportion of entrepreneurs in our sample (TEA=2.9%). The top layer split in the model is the working status of individuals. If people currently hold a job (part-time or full-time), their chances of becoming entrepreneurs increase to 5.2%. Otherwise it drops to 0.4%. Among those that are currently employed, the SUSKILL variable turns out to be the only significant predictor able to differentiate between entrepreneurs and non-entrepreneurs in the tree model. Those showing confidence in their ability exhibit a chance of 15.8% to become an entrepreneur, otherwise chances are only 3.3%

Table 3 summarizes CART results across all our 18 trees. Overall, the perception of having sufficient skills to start a new business is the most influential variable in the CARTs computed for the 18 countries.

Table 3 about here

In 15 of the 18 trees, SUSKILL is the best split in the first layer of the tree. That is, it is the single best predictor with the highest classification performance regarding the dependent variable. For three countries (Japan, South Korea, New Zealand), working status yields the best first-layer split. In these three trees, however, the perception of sufficient skills appears as best split in the second layer of the tree. Thus, overall, the CARTs indicate the perception of sufficient skills as the main driver of an individual's decision to start a new business across all 18 countries in our sample, even when close attention is paid to differences across countries.

5 Discussion

Overall, our results suggest that education and income do not play an important role in an individual's decision to become an entrepreneur since neither variable appears as a significant node in any of the CART trees. Our results, however, do lend support to existing empirical evidence suggesting that the relative impact of a crucial set of variables on the rates of new business creation is significantly influenced by geographic and historical circumstances (Acs and Evans 1994, Blanchflower 2004, Reynolds et al. 2003). We also test whether individuals with a high propensity to become entrepreneurs exhibit the same characteristics in different countries. The comparison of CARTs clearly shows that entrepreneurs in various nations possess similarities, but also remarkable differences. Among the similarities, the most striking is the importance, across all trees, of the sufficient skill variable. That is, the subjective perception that individuals have of their own entrepreneurial abilities.

According to the findings, the sufficient skill variable (suskill) and work status (gemwork) are the more prominent features of a high propensity to be an entrepreneur. The perception of sufficient skills is clearly the dominant variable in the CART analysis. In 18 final trees derived for the countries in our sample, 15 had the sufficient skill variable at the top (83%). In Italy, Portugal, and Finland, the sufficient skills variable provides the only significant split. Three trees had the work status at the top (South Korea, Japan, New Zealand). But in all of them, sufficient skills appeared in the second layer of the tree. Hence the perception of sufficient skills is a dominant variable that seems to have an effect independently from institutional settings, culture and overall level of entrepreneurial activity.

"Knowing an entrepreneur" is also somewhat important, but the strength of this effect differs significantly across countries. The positive impact of knowing an entrepreneur might be explained via the effect of role models and networks or as a reduction in perceived risk. Because most individuals are risk averse and risk aversion is applied to perceived rather than objective risks, a reduced risk perception should increase the probability of an individual to start a new business (Weber and Milliman 1997). Also, as mentioned earlier, the education variable never appears to play a role in the decision to start a new business. Education is often used as an "objective" measure of potential entrepreneurial skills. To the extent that education is indeed a good proxy, the lack of significance of the interaction term suggests the lack of any correlation between individuals' perceptions and actual abilities.

To summarize, consistently with existing literature (Arenius and Minniti forthcoming, Koellinger et al. 2004), "having sufficient skills" appears to be the main driver of the decision to start a new business. This variable is clearly a perceptual variable and, as a result, is likely to be biased. Indeed, there is some evidence that distortions in perceptions are relatively common among potential entrepreneurs (Busenitz and Barney, 1997; Cooper et al., 1988). The importance of perceptual variables, and their associated bias, in the decision to start a new business, may explain some of the observable inconsistencies between returns to entrepreneurship and entrepreneurial decisions found in the literature.

It is a well known fact that many new businesses fail shortly after inception (Baldwin, 1995; Dunne et al., 1988). Camerer and Lovallo (1999) hypothesize that the high rate of business failure may be, in part, the result of managers acting on optimism about their relative skills. Marketing managers, for example, have been shown to be too confident with respect to their judgment, and that their perceptional bias is more pronounced if they are more experienced (Mahajan 1992). Busenitz and Barney (1997), however, have shown that entrepreneurs' overconfidence is even higher than the overconfidence of managers. Finally, Cooper et al. (1988) reported on even more striking findings regarding entrepreneurs' overconfidence. In a study of 2,994 entrepreneurs, 81% were shown to believe that their success chances were at least 70%, while a third believed that their success was certain. Moreover, most of them believed that their chances of survival were higher than those of competitors. Unfortunately, the reality at this time was that 66% of all newly founded enterprises failed.

Our results suggest that the optimistic bias caused by overconfidence in one's own skills is an important determinant of an individual's decision to start a new business. If potential entrepreneurs tend to be overconfident about their relative skills, then we should expect a large number of new business failures and relatively low returns to entrepreneurial activity. This is consistent with Hamilton (2000), who found entrepreneurship to be a career choice that does not pay. In fact, he shows that except for the dynamic comparison of the highest 25% of entrepreneurial and wage incomes, staying in a wage job or moving back to it makes more economic sense than to remain self-employed. Our results are also consistent with Moskovitz and Vissing-Jørgensen (2002) who investigated the risk-return relationship of private enterprises and found strong evidence identifying them as inferior investments. According to them, the "private equity premium puzzle" might be explained by: "high entrepreneur risk tolerance, large additional pecuniary benefits, large non pecuniary benefits, a preference for skewness, and over optimism and misperceived risk" (2002, p. 747).

6 Implications for future research

In this paper, we use an original data set and test for the relative importance of socio-economic and perceptual factors on the decision to start a new business. Individuals who perceive their skills as sufficient to starting a new business are shown to be more likely to do so. In addition, knowing another entrepreneur is shown to increase further the likelihood of starting a new business. Likely, this is the case because the perception of the risk associated with owning one's own business decreases as other entrepreneurs, whether or not successful, may be observed.

Clearly, our results are suggestive and more work in the area is required. In principle, an individual's perception of skills could be based on objective skills not captured in our data set. Moreover, our data set may not include all the relevant variables and may not capture the true direction of the causal relationship between dependent and independent variables. Our CART analysis, for example, suggests the possibility of unaccounted factors embedded in a country's environment. Among the unaccounted factors, the institutional framework is likely to be an important omission. Thus, further extensions of this project may include proxies for institutions. Unfortunately, the data available for 2001 do not allow us to do so.

The institutional framework is crucial in determining the quantity and quality of entrepreneurial behavior as it defines individuals' incentives to transform perceived opportunities into actions. In the long run, the institutional framework mold individuals' perceptions. Harper (1998) argues explicitly that the nature of political and economic institutions influences individuals' perceptions. Those institutions and policies that improve transparency and entitlement tend to increase the subjective perception of the link between actions and outcome. They increase, therefore, the number of individuals who perceive themselves as having an internal locus of control. Along similar lines, Baumol (1990) argues that institutional arrangements affect the quantity and type of entrepreneurial efforts. An institutional setting leading to stronger perceptions of control over one's domain should yield more entrepreneurial activity. It is our hope that our work will spur interest and much needed research on the relationship between institution and entrepreneurship.

Finally, individual perceptions may differ from actual abilities and risk level. On the other hand, a person might perceive her own "entrepreneurial alertness" as a signal of potential success, and such an assumption could be appropriate. As a result, although the entrepreneurial environment may be crowded with individuals acting on overconfident self-perceptions, this is not to say that potential entrepreneurs behave irrationally. They are simply overconfident and the "inside view" leads them, often, to overestimate their own skills. In the long run, because of its push toward taking initiative even in spite of unfavorable odds, overconfidence may emerge as an individual's winning strategy.

References

Acs, Z., D. Evans. 1994. The determinants of variations in self-employment rates across countries and over time. Working paper.

Arenius, P., M. Minniti. Forthcoming. Perceptual Variables and Nascent Entrepreneurship. Small Business Economics

Baldwin, J. R. 1995. The dynamics of industrial competition. Cambridge: Cambridge University press.

Bates, T. 1990. Entrepreneur human capital inputs and small business longevity. Review of Economics and Statistics 72: 551–559.

Baumol, W. J. 1983. Toward operational models of entrepreneurship. In: J. Ronen (Ed.), Entrepreneurship. Lexington, MA: Lexington Books.

Baumol, W. J. 1990. Entrepreneurship: Productive, unproductive, and destructive. Journal of Political Economy 98: 893–921.

Baumol, W. J. 1993. Formal entrepreneurship theory in economics: Existence and bounds. Journal of Business Venturing 8 ; 197–210.

Blanchflower, D.G. 2004. Self-Employment: More may not be better. NBER Working Paper No. 10286

Blanchflower, D.G., A.J. Oswald. 1990. Self-employment and the enterprise culture. In British Social Attitudes: the 1990 Report. Jowell, R. and S. Witherspoon (eds.). Gower Press, Aldershot.

Blanchflower, D.G., A.J. Oswald. 1998. What makes an entrepreneur? Journal of Labor Economics 16(1): 26-60.

Bogenhold, D., U. Staber. 1991. The decline and rise of self-employment. Employment and Society 5: 223-239.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.L. 1984. Classification and Regression Trees. Statistics/Probability Series. Belmont (CA-USA): Wadsworth International Group.

Busenitz, L. W., J. B. Barney. 1997. Differences between Entrepreneurs and Managers in Large Organizations: Biases and Heuristics and Strategic Decision-Making. Journal of Business Venturing 12: 9-30.

Camerer, C., D. Lovallo. 1999. Overconfidence and Excess Entry: An Experimental Approach. American Economic Review 89: 306-318.

Chell, E., S. Baines. 2000. Networking, entrepreneurship and microbusiness behavior. Entrepreneurship and Regional Development 12: 195–215.

Cooper, A. C., C.Y. Woo, and W. C. Dunkelberg. 1988. Entrepreneurs' Perceived Chances for Success. Journal of Business Venturing 3: 97-108.

Dunne, T., M. J. Roberts, and L. Samuelson. 1988. Patterns of firm entry and exit in U.S. manufacturing industries. Rand Journal of Economics 19(4): 495-515.

Evans, D., B. Jovanovic. 1989. An estimated model of entrepreneurial choice under liquidity constraints. Journal of Political Economy 97: 808–827.

Evans, D., L. Leighton. 1989. Some empirical aspects of entrepreneurship. American Economic Review 79: 519–535.

Gatnar, E. 2002. Tree-based models in statistics: Three decades of research. in: Jajuga, K., Sokolowski, A. and Bock, H.-H. (eds). Classification, Clustering, and Data Analysis. Berlin et. al.: Springer. 399-407.

Hamilton, B. H. 2000. Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment. Journal of Political Economy 108: 604-631.

Harper, D. 1998. Institutional Conditions for Entrepreneurship. Advances in Austrian Economics 5: 241-275.

Hayek, F.A. 1952. The Sensory Order. Chicago: The University of Chicago Press.

Holtz-Eakin, D., D. Joulfaian, and H.S. Rosen. 1994a. Entrepreneurial decisions and liquidity constraints. Journal of Political Economy 102: 53-75.

Holtz-Eakin, D., D. Joulfaian, and H.S. Rosen. 1994b. Sticking it out: entrepreneurial survival and liquidity constraints. Rand Journal of Economics 25(2): 334-347.

Iyigun, M., A. Owen. 1998. Risk, entrepreneurship and human capital accumulation. American Economic Review 88: 454–457.

Kahneman, D., D. Lovallo. 1993. Timid choices and bold forecasts: A cognitive perspective on risk taking. Management Science 39(1): 17-31.

Kihlstrom, R., J. Laffont. 1979. A general equilibrium entrepreneurial theory of firm formation based on risk aversion. Journal of Political Economy 87: 719–740.

Kirzner, I. M. 1973. Competition and entrepreneurship. Chicago, IL: University of Chicago Press.

Kirzner, I. M. 1979. Perception, opportunity, and profit. Chicago, IL: University of Chicago Press.

Koellinger, P., M. Minniti, and C. Schade, 2004, I think I can, I think I can: A cross-country study of entrepreneurial motivation. Working paper. Humboldt University, Germany.

Lévesque, M., M. Minniti. Forthcoming. The effect of aging on entrepreneurial behavior. Journal of Business Venturing.

Mahajan, J. 1992. The Overconfidence Effect in Marketing Management Predictions. Journal of Marketing Research 39: 329-342.

Moskowitz, T. J., A. Vissing-Jørgensen. 2002. The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle? American Economic Review 92: 745-778.

Otani, K. 1996. A human capital approach to entrepreneurial capacity. Economica 63: 273–289.

Reynolds, P., E. Autio, and M. Hay. 2003. Global Entrepreneurship Monitor Report. Kansas City, MO: E. M. Kauffmann Foundation.

Reynolds, P.D., N. Bosma. 2005. Cross-National Entrepreneurial Assessments: The GEM Operational Procedures Small Business Economics

Sheskin, D.J. 2000. Handbook of Parametric and Nonparametric Statistical Procedures. 2nd edition, Boca Raton et. al. : Chapman & Hall.

Taylor, M.P. 1996. Earnings, independence or unemployment; why become self-employed? Oxford Bulletin of Economics and Statistics 58(2): 253-265.

Weber, E. U., R. A. Milliman. 1997. Perceived Risk Attitudes: Relating Risk Perception to Risky Choice. Management Science 43: 123-144.

Zhang, H., M. Bracken. 1995. Tree-based risk factor analysis of pre-term delivery and small-for-gestational-age birth. American Journal of Epidemiology 141: 70-78.

Zhang, H., B. Singer. 1999. Recursive Partitioning in the Health Sciences, New York et. al.: Springer.

Appendix – CART Method

Splitting nodes

In CART, the sample of subjects is systematically sorted into completely homogeneous subsets until a saturated tree is found. For each split, CART considers the entire set of available predictor variables to determine which one maximizes the homogeneity of the following two daughter nodes. This is a hierarchical process that reveals interdependencies between covariates. The process is continued until the nodes are completely homogeneous and cannot be split any further. Breiman et al. describe a number of possible splitting methods (Breiman et al., 1984, ch. 4). Among them, the entropy impurity criterion is identified as the best method for the identification of the predictors of a dependent variable with low frequency. Consider the splitting of a parent node, where *a, b, c,* and *d* denote the number of subjects in the two daughter nodes:

	Predictor	Entrepreneur	Non-Entrepreneur	
Left node (t_L)	$s_i = 1$	A	B	a+b
Right node (t_R)	$s_i = 0$	C	D	c+d
		a+c	B+d	n = a+b+c+d

From Breiman et. al. (1984, pp. 94-102), the entropy impurity in the left daughter node is

$$(1)\ i(t_L) = -\frac{a}{a+b}\log\left(\frac{a}{a+b}\right) - \frac{b}{a+b}\log\left(\frac{b}{a+b}\right).$$

Similarly, the entropy impurity in the right daughter node is

$$(2)\ i(t_R) = -\frac{c}{c+d}\log\left(\frac{c}{c+d}\right) - \frac{d}{c+d}\log\left(\frac{d}{c+d}\right).$$

Consequently, the impurity of the parent node is

$$(3)\, i(t) = -\frac{a+c}{n}\log\left(\frac{a+c}{n}\right) - \frac{b+d}{n}\log\left(\frac{b+d}{n}\right).$$

The goodness of a split, *s*, is then measured by

$$(4)\ \Delta I(s,t) = i(t) - P\{t_L\}i(t_L) - P\{t_R\}i\{t_R\}$$

Where P{t} is the probability associated with the occurrence of the each daughter node. The goodness of a split is calculated for all available predictor variables. The split characterized by the highest $\Delta I(s,t)$ allows the identification of the best predictor. This recursive partitioning process continues until the tree is saturated. That is, nodes cannot be split any further because the subjects they contain are perfectly homogeneous. T_0 is the saturated tree. The saturated tree is usually too large to be useful. And, in the worst case, it is trivial because each terminal node could consist of just one case. Of course, the resulting model is also subject to severe over-fitting problems. As a result, it is necessary to find a nested sub-tree of the saturated tree that exhibits the best "true" classification performance and satisfies statistical inference measures.

Pruning

The purpose of pruning is to find the right-sized tree, which should be a sub-tree of T_0. We use the cost-complexity pruning algorithm suggested by Breiman et. al. (1984), which ensures that a unique best sub-tree can be found for any given tree complexity. The right sized tree should not be subject to over-fitting and insignificant splits, but detailed enough to exhibit a good classification performance. Recall that CART predicts the outcome (e.g. entrepreneur or non-entrepreneur) based on the group membership of a case in the sample. In the tree, each subject falls into exactly one terminal node. We choose a class assignment rule that assigns a class to every terminal node $t \in \widetilde{T}$. In our application, node *t* is assigned "entrepreneur

{Y=1}" if $P\{Y=1|t\} \geq 0.5$ and vice versa. In this simple case, the expected cost resulting from any subject within a node is given by

(5) $r(t) = 1 - P(i \mid t)$,

where $P(i \mid t)$ is the percentage of misclassified subjects in a node.[1] The classification performance of the entire tree is given by the quality of its terminal nodes

(6) $R(\mathrm{T}) = \sum_{t \in \tilde{T}} P(t) r(t)$,

where $R(\mathrm{T})$ is the misclassification cost of all terminal nodes in the tree, $\tilde{\mathrm{T}}$ the set of terminal nodes, and $P(t)$ the probability of a subject to fall into the terminal node t.

We are now ready to turn to the main idea of cost-complexity pruning (Breiman et. al., 1984, pp. 66-71): For any subtree $\mathrm{T} \leq \mathrm{T}_0$, define its complexity as $|\tilde{\mathrm{T}}|$, the number of terminal nodes in T. Let $\alpha(\geq 0)$ be a real number called the complexity parameter and define the cost complexity of the entire tree as

(7) $R_\alpha(\mathrm{T}) = R(\mathrm{T}) + \alpha|\tilde{\mathrm{T}}|$.

For any value of $\alpha(\geq 0)$, there is a unique smallest subtree of T_0 that minimizes $R_\alpha(\mathrm{T})$. The formal proof is in Breiman et. al. (1985, chapter 10). Thus, by gradually increasing α, a sequence of nested essential subtrees of T_0 can be constructed by pruning off the weakest branches at each threshold level of α. Note that T_0 minimizes $R_\alpha(\mathrm{T})$ if $\alpha = 0$. If α becomes large enough, the root node becomes the optimal solution.

[1] Note that *r(t)* becomes smaller for any additional split. Thus, *r(t)* is minimal for the saturated tree. See Breiman et. al. (1984, p. 95-96) for a formal proof.

Selection of the best pruned tree using cross-validation

The classification performance $R(\mathrm{T})$ as specified in (6) is obviously biased and results in severe over-fitting. To select the best pruned tree, we need a more honest estimate of the true misclassification cost of the tree. This is usually done with an independent test sample, e.g., boot-strapping or cross-validation. However, we choose a 20-fold cross validation procedure because it makes better use of the information contained in the original dataset than the independent test sample method and, in addition, it outperforms bootstrapping in terms of reduced bias (Breiman et. al., 1984, pp. 72-78, 311-313). We estimate $\hat{R}(\mathrm{T})$ by growing a series of V auxiliary trees together with the main tree grown on the learning sample Λ. The V auxiliary trees are grown on randomly divided, same sized subsets, Λ_V, $v = 1,...,V$, with the v-th learning sample being $\Lambda^{(v)} = \Lambda - \Lambda_v$ so that $\Lambda^{(v)}$ contains the fraction $(V-1)/V$ of the total data cases. For each v, the trees and their pruning sequence are constructed without ever seeing the cases in Λ_V. Thus, they can serve as an independent test sample for the tree $T^{(V)}(\alpha)$. The idea now is that for V large, $T^{(V)}(\alpha)$ should have about the same classification accuracy as $T(\alpha)$. The estimated misclassification costs $\hat{R}(\mathrm{T})$ equal the proportion of misclassified test set cases in the V auxiliary trees at the α complexity levels. The best pruned tree is the one with the smallest $\hat{R}(\mathrm{T})$.

Significance of splits

Finally, the significance of each individual split in the selected tree can be tested following Sheskin (2000; section 16.6). Recall that we calculate the re-substitution risk as

$$(8)\quad r = \frac{a}{a+b} \Big/ \frac{c}{c+d}$$

The calculation of the confidence interval of *r* requires to compute the standard error of the two daughter nodes, which is given by

$$(9)\quad SE_r = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}\,.$$

Since the sampling distribution of the re-substitution risk is positively skewed, a logarithmic scale transformation is employed in computing the confidence interval (Christensen, 1990; Pagano and Gauvreau, 1993). The α -confidence level is obtained by

$$(10)\quad \left\{ e^{\left[\ln(r) - SE \bullet z_\alpha\right]}; e^{\left[\ln(r) + SE \bullet z_\alpha\right]} \right\},$$

where z_α is the tabled two-tailed *z* value for the $(1-\alpha)$ confidence level. For the 95% confidence level, the relevant .05 value is $z_{.05} = 1.96$. This test is computed for all splits in the tree that was selected from the pruning sequence after the cross-validation procedure.

Table 1: Un-weighted TEA 2001 ratios in 18 countries

Country	**Total entrepreneurial activity 2001**	**N**
AR - Argentina	9.6%	1,992
CA – Canada	9.1%	1,939
D – Germany	5.8%	7,058
DK – Denmark	5.6%	2,022
FIN – Finland	5.1%	2,001
HU – Hungary	10.9%	2,000
IN – India	11.7%	2,011
IL – Israel	3.8%	2,055
IT – Italy	8.2%	1,973
JP – Japan	2.9%	2,000
KR – South Korea	13.4%	2,008
NZ – New Zealand	15.1%	1,960
P – Portugal	6.6%	2,000
PL – Poland	7.1%	2,000
RU – Russia	6.0%	2,012
S – Sweden	4.9%	2,056
SG – Singapore	5.9%	2,004
US – United States	9.1%	2,954
TOTAL	7.6%	42,045

Source: GEM 2001 survey data

Table 2: Variable definitions and un-weighted descriptive statistics, GEM 2001 data

Variable (corresponding survey question)	Value	Relative Frequency
Gender	Male	48.1%
	Female	51.9%
Knowent (You know someone personally who started a business in the past 12 months.)	Yes	33.6%
	No	64.8%
	Refused	1.6%
Opport (In the next six months there will be good opportunities for starting a business in the area where you live.)	Yes	23%
	No	60.6%
	Refused	15.8%
Suskill (You have the knowledge, skill and experience required to start a new business.)	Yes	36.3%
	No	58.7%
	Refused	4.9%
Fearfail (Fear of failure would prevent you from starting a new business.)	Yes	33.2%
	No	60%
	Refused	6.7%
Famfutur (Looking ahead, do you think that a year from now you and your family will be better off financially, or worse off, or about the same as now?)	Worse	14.4%
	Same	49.1%
	Better	29.2%
	Missing	7.2%
Ctrfutur (In a year from now, do you expect that in the country as a whole business conditions will be better or worse than they are at the present, or just about the same?)	Worse	25%
	Same	38.2%
	Better	24.6%
	Missing	12.2%
Gemwork (Present working status of the individual)	Full / Full or part time	50.3%
	Part time only	6.9%
	Retired / disabled	10%
	Homemaker	10%
	Student	5.1%
	Not working: other	17%
	Missing	0.7%
Gemhhinc (Household income of the individual recoded into thirds relative to country income distribution.)	Lowest 33%	26.4%
	Middle 33%	30.9%
	Upper 33%	20.9%
	Missing	21.8%
Gemeduc (Educational attainment of the individual.)	Some second. school..	26.9%
	Secondary degree	34.9%
	Post second. degree	33.3%
	Grad exp	1.4%
	Missing	3.5%
Age – in 8 categories (What year were you born?)	14-17 yrs old	2.1%
	18-24 yrs old	13%
	24-34 yrs old	19.2%
	35-44 yrs old	21.5%
	45-54 yrs old	18.1%
	55-64 yrs old	14.5%
	65-74 yrs old	8.1%
	75-84 yrs old	3%
	85-up yrs old	0.4%

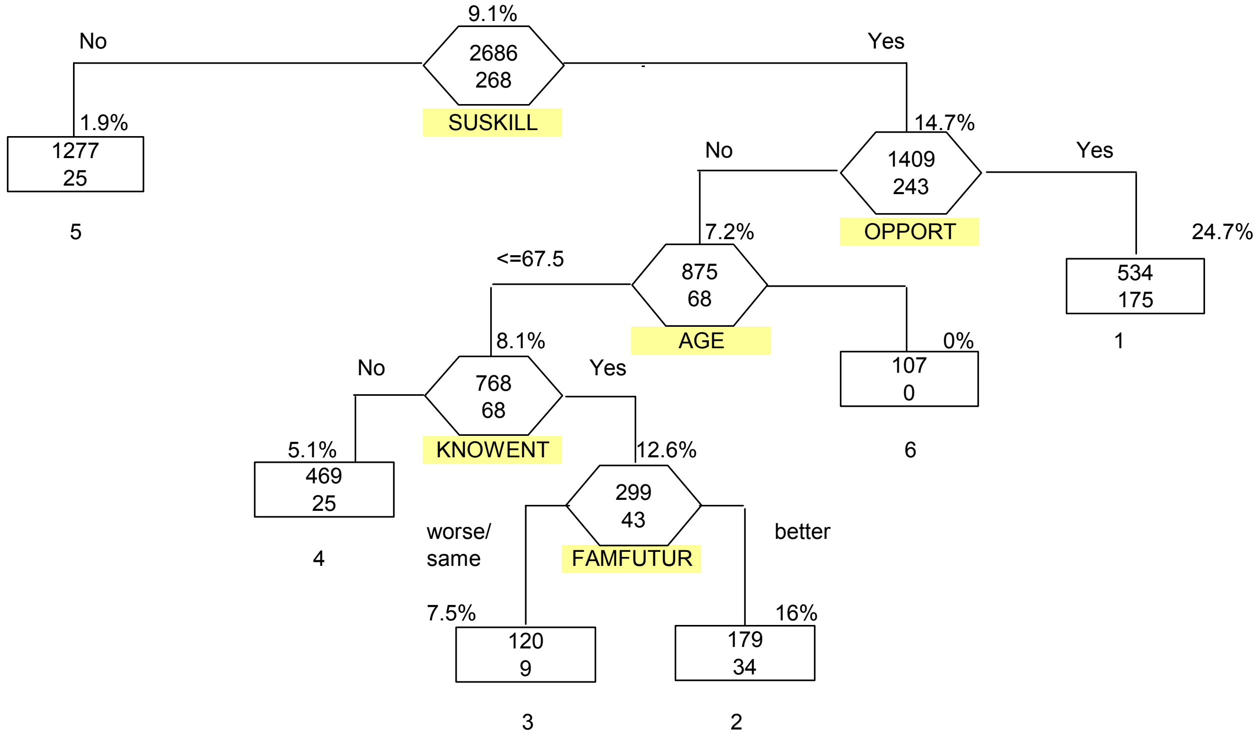
Base: AR, CA, D, DK, FIN, HU, IN, IL, IT, JP, KR, NZ, P, PL, RU, S, SG, US. *N = 42,045*

Table 3: *Occurrence of predictor variables in top layers of 18 trees*

Variable	**Absolute occurrence as 1st layer split**	**Absolute occurrence as 2nd layer split**
Sufficient skills	15	3
Working status	3	6
Opportunities		4
Knowing an entrepreneur		2
Age		1
Fear of failure		1

Base: 18 CARTs on TEA01 in AR, CA, D, DK, FIN, HU, IN, IL, IT, JP, KR, NZ, P, PL, RU, S, SG, US

__Figure 1:__ CART for total entrepreneurial activity in the USA 2001 (tea01)



Note: US-CART structure remains "complex" if a randomly drawn sub-sample with 2,000 observations is used.

Figure 2: *CART for total entrepreneurial activity in Singapore 2001 (tea01)*

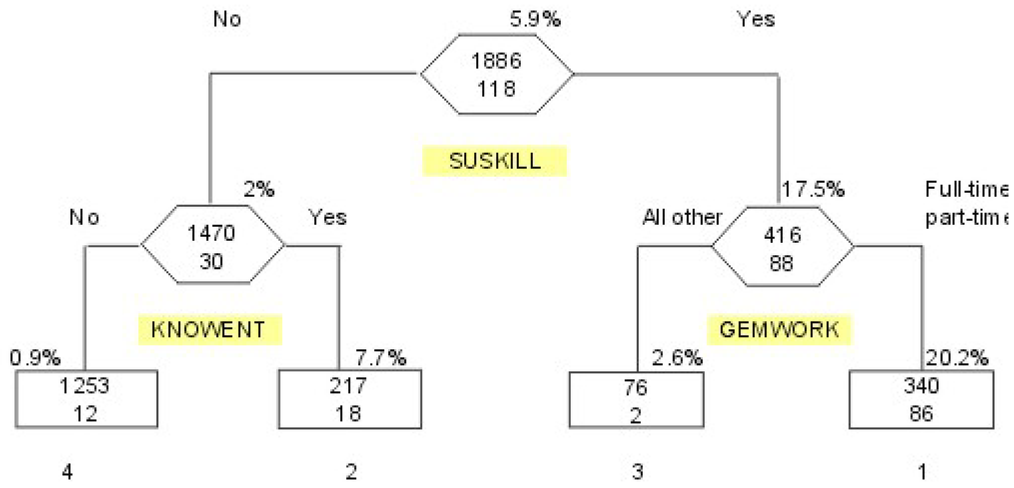


Figure 3: CART for total entrepreneurial activity in Denmark 2001 (tea01)

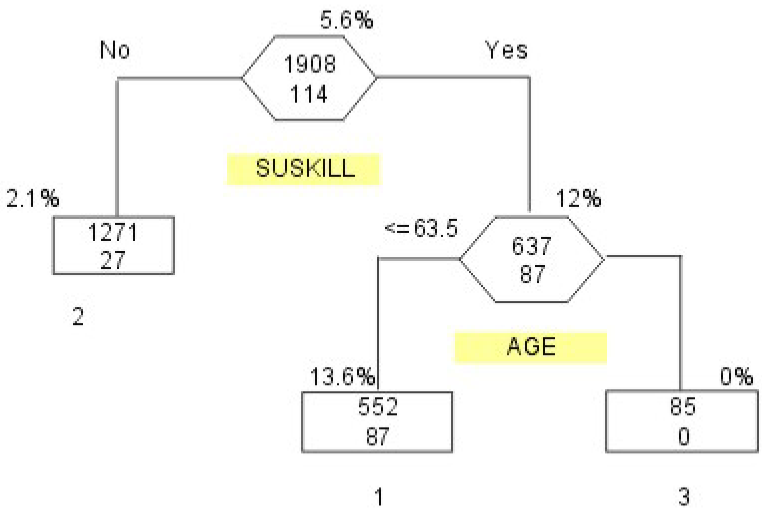


Figure 4: CART for total entrepreneurial activity in Japan 2001 (tea01)