ASSESSING THE Commercial Viability OF NEW VENTURES

Research shows VCs have a poor track record when it comes to picking a winner. Can a statistical model help?

BY THOMAS ÅSTEBRO

Only a small fraction of all start-ups receive venture capital (VC). That's because while VC firms argue that there are few high-quality start-ups to invest in, Canadian entrepreneurs claim that there is a lack of venture capital for start-ups. The VC industry cites a lack of reliable information as a reason to shy away from start-ups. So, to help with venture capital decision-making, I developed a statistical model that predicts the likelihood of a new venture successfully reaching the market. This model is based on data from over 500 new Canadian ventures and successfully predicts correct market outcomes in 83 per cent of the cases. It is significantly better at predicting the outcomes of new ventures than seasoned venture capitalists (VCs) whose predictions have an accuracy ranging between 17 and 40 per cent. This model can be used to screen venture capital applicants.

Oversupply of Low-Quality Ventures or Undersupply of Capital?

The Canadian venture capital industry has often been criticized for not investing enough in new Canadian ventures. One explanation given by VCs for avoiding start-ups is their inability to make informed decisions for these ventures. For example, there is typically a lack of information about the venture's market potential. Most financial and market information takes the form of forecasts which all exhibit the typical "hockey-stick" growth curve. Entrepreneurs are not to blame for making such forecasts because they are coached by advisors to produce them for VCs, who themselves claim not to be interested in ventures without "significant potential." But how does one assess what constitutes a significant market potential? And how does one assess other important venture characteristics such as the uniqueness of the underlying technology, the threat of competitive responses from established firms and the completeness and ability of the management team? Ultimately, how does one know which factors are important and which factors aren't?

Venture Capital Decision-Making

There are three potential decision-making methods. The first is normative, based primarily on operations research principles. However, the complexity of this approach doesn't make it of much use for VC decision-making. The second option is judgmental. The typical judgmental model used by VCs is the "gut feel" approach. A version of a judgmental review uses expert systems ("checklists") that encode the decision-making criteria and rules of an expert, or group of experts. But if the encoded expert is wrong in his/her gut feel, then the expert system will be equally wrong. A third option used is a statistical model based on historical data.

There is a large body of research on the judgmental factors venture capitalists use when making investment decisions. Zopounidis (1994) summarizes the literature with two conclusions: "The first is that the criterion of the management team is considered predominant in all the studies concerning decisions in venture investment and the second is the great diversity of evaluation criteria and their relative importance (ranking of criteria) from one study to the other" (63).

A recent Ph.D. thesis by Jagdeep Bacchher at the University of Waterloo investigated the judgment factors deemed important by leading U.S. and U.K. VCs assessing seed and early-stage technology-based ventures (Bacchher, 2000). Most evaluations were of dot-com

Thomas Åstebro is associate professor, of management sciences at the University of Waterloo.

ventures. Again, a multitude of factors were used, the most important being management capabilities, market opportunity, and return on investment. A problem with these studies is that the decision-making criteria have not been successfully calibrated against actual outcomes. That makes it difficult to know whether the factors VCs believe to be important actually are.

Andrew Zacharakis and Dale Meyer (2000) investigated the ability of VCs to accurately assess the future success of a venture-seeking investment. The 51 practising U.S. VCs interviewed had, on average, over ten years of VC experience and over 22 years of work experience focused on seed and early-stage deals. They conducted an experiment where the VCs received several pieces of information about 25 actual investments that had subsequently achieved either success or failure. The VCs were requested to evaluate the ventures as they would during the initial screening stage. Approximately 57 per cent of the investments were in seed and early stage. Their predictions were then compared to the actual outcomes and a "hit-rate" was computed — the percentage of correctly classified outcomes.

A rather surprising result of this study was the low ability of the VCs to correctly forecast the outcomes of the ventures — at best the VCs had a hit-rate of approximately 40 per cent. Perplexingly, the more information about the venture that was provided to the VCs, the less able they were to correctly predict outcomes. When information about the track record of the team and competition was included, the hit-rate was reduced to 31 per cent and when additional information about the team and product was introduced, the hit-rate declined to 17 per cent. These results indicate that VCs are rather poor at making investment decisions and that more information makes them more confused and less accurate evaluators.

Zacharakis and Meyer then compared the ability of the VCs to the forecasting accuracy of a simple statistical model that used the same information as the VCs to make predictions. The hit-rate of the statistical models ranged from 40 per cent to 60 per cent, always clearly surpassing the judgments made by the VCs.[2]

"The more information about the venture that was provided to the VCs, the less able they were to correctly predict outcomes."

Why Humans are Worse Decision-Makers than Statistical Models

Venture capital decision-making is not the only area in which humans have been found to be inferior classifiers compared to statistical models. In fact, the main conclusion from several hundred studies on judgmental versus statistical decision-making models is that statistical models are at least equal and mostly superior to judgmental decision-making (Dawes et al., 1989; Grove and Meehl, 1996). This conclusion holds true across a number of decision-making contexts, both real and experimental, expert and novice.

A number of objections have been raised to these conclusions. One is that judgment mediated by theories is superior to statistical (essentially theory-less) analysis. A slight advantage has indeed been found in the medical sciences for clinical judgment resting on firm theoretical grounds (Dawes et al., 1989). However, this cannot be said about the social sciences, where theories are typically rather unsuccessful at predicting outcomes. A second objection is that experts might gain advantage by recognizing rare events (the "broken leg" cue) or complex patterns. However, when experts are provided with both the available data and the statistical prediction to search for these exception cues, experts typically perform no better or add very little to the statistical prediction (Dawes et al., 1989; Einhorn, 1972). One argument for this result is that experts tend to ascribe too much weight to exceptions. In a review of this particular issue, Bunn and Wright (1991) compile some evidence suggesting that experts, while not competitive with statistical models, can augment predictions of statistical models, particularly for time-series and weather forecasting. And it is also recognized that humans have a superior capability in visual pattern recognition such as facial expressions, in language translation and for inventing deep-structure theories.

There are several reasons for the common failure of humans over statistical models. Humans have difficulties processing large amounts of data in parallel and distinguishing valid and invalid variables. They also have problems dealing with sample selection bias and data truncation. Humans also have a tendency to let judgment be affected by recent events, by hindsight bias, and tend to seek only confirmatory data. They also have the tendency to be over-confident.[3]

My conclusion is thus that institutional investors are likely to have difficulties similar to those of venture capitalists in judgmentally assessing the future success of their investments. In addition, investors that speculate on the public market face the problem of competing with automatic trading programs. Not only can such investors be disadvantaged by decision-making biases, but also by their slower speed of reaching decisions compared to statistical models. On the other hand, there are a large number of automatic trading programs operating in the public equity market that, over the long run, are likely to nullify any slight advantage a specific trading program might have (Sullivan et al., 1999). My conclusion is that it is unlikely that humans have any specific advantage in the public equity market unless they have insider (private) information. It therefore remains for them to play in the private equity market and take advantage of the relative scarceness of public information to form superior investment strategies based on private information.

Predicting the Commercial Viability of New Ventures

I investigated the ability of both a statistical model and experts to forecast the probability that an early-stage venture will be commercialized. I sampled a group of 561 new ventures that were submitted by entrepreneurs for commercial evaluation during the period between 1989 and 1993 to the Inventor's Assistance Programs (IAP) at the Canadian Innovation Centre in Waterloo, Ontario. Skilled engineers that had some business knowledge evaluated the ventures. The ventures were evaluated at a very early stage of development, having low initial research and development (R&D) efforts and none yet being commercialized. I compared the ex ante evaluations and forecasts by experts at the IAP with the ex post commercial success of these 561 ventures. Data on a broad range of project characteristics were evaluated by the IAP, spanning technology, market, distribution, business logic, legal, production and risk factors. The expert subjectively rates the project on 37 criteria and, based on an intuitive judgment of the combination of scores on those criteria, determines an overall score for the project. Since the method of assessing the joint effect of the criteria is judgmental, the overall assessment might differ across evaluations and evaluators, even though data are identical. The overall score was easily converted into a forecast of success or failure. These forecasts were compared to objective data on the ventures' outcomes in terms of commercial success (or failure) that were collected in 1996 through a telephone survey to the entrepreneurs.

> "Humans also have a tendency to let judgement be affected by recent events and tend to seek only confirmatory data."

Tests showed that these experts correctly predicted no less than 79 per cent of the outcomes correctly. In the last year of evaluation the experts correctly predicted 83.8 per cent of the outcomes correctly.

I then devised a statistical model using the data on the 37 criteria and the observed outcomes. This model contained only four criteria that were statistically significant: "Expected Profitability," "Development risk," "Functional Performance" and "IP Protection." These were defined as questions:

• Will the expected revenue from the innovation provide more profits than other investment opportunities?
• What degree of uncertainty is associated with complete, successful development from the present condition of the innovation to the market-ready state?
• Does this innovation work better than the alternatives?
• Is it likely that worthwhile commercial protection will be obtainable for this innovation through patents, trade secrets or other means?

The statistical model correctly predicts 82.6 per cent of all outcomes in forward cross-validation tests. These tests mean that the model was developed on a specific sample for a given time period and then tested for its predictive accuracy on a different sample from a different time period.

The Results

Table 1 displays the results. In 1993 the IAP experts correctly predicted 12 successes (70.6 per cent) and 128 failures (85.3 per cent) for an overall prediction accuracy of 83.8 per cent (see columns 2 and 3). The statistical model developed on the 1989 to 1992 pool of data correctly predicted 11 successes (64.7 per cent) and 127 failures (84.7 per cent) for an overall forward prediction accuracy of 82.6 per cent. These data are displayed in columns 4 and 5. In this comparison, the experts gain a slim victory over the statistical model. But the number of observations that make up the difference is so small (two) that one can hardly claim the experts are outperforming the statistical model. I tried to improve on the four-variable model in different ways.

> "The statistical model correctly predicts 82.6 per cent of all outcomes in forward cross-validation tests."

One approach is to include all 37 variables. Not surprisingly, by using all 37 predictors to fit the 1989 to 1992 data, we find the best model overall. However, when applying this model to the 1993 sample it had an overall prediction accuracy of merely 71.9 per cent (see columns 6 and 7). These are clear indications of model over-fitting, especially in the model's tendency to (falsely) overestimate the number of successes. Another approach is to include predictors with higher *p*-values, but not all 37. I explored this using stepwise regression with an inclusion criterion of $p < 0.10$. In addition to the four predictors this allows entry of five more variables for the 1989 to 1992 pool. Applied to the 1993 pool, the model correctly predicts 12 successes (70.6 per cent) and 122 failures (81.3 per cent) for an overall forward prediction accuracy of 80.2 per cent (see columns 8 and 9). The new model does not improve prediction accuracy. The reason appears to be the same as that of the model with all 37 variables, a tendency to (falsely) overestimate the number of successes. Another potential improvement is to specify interaction effects among the significant predictors. Including all two-way interactions does not, however, improve out-of-sample predictions (see columns 10 and 11).

Conclusion

The judgmental process used by experts at the IAP to make an overall assessment of an early-stage venture's commercialization prospects is extremely accurate with at least twice the hit-rate of seasoned VCs in the U.S. The IAP experts are approximately just as

Table 1.

FORWARD CROSS-VALIDATION OF PREDICTIVE ACCURACY.

	IAP		STATISTICAL MODEL**							
	Overall Rating*		*p*<0.05		37 Variables		*p*<0.10		2-Way Interactions	
	No.	%	No.	%	No.	%	No.	%	No.	%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Overall Predictive Accuracy	140	83.8%	138	82.6%	120	71.9%	134	80.2%	138	82.6%
Correctly Predicts Success (Sensitivity)	12	70.6%	11	64.7%	12	70.6%	12	70.6%	11	64.7%
Correctly Predicts Failure (Specificity)	128	85.3%	127	84.7%	108	72.0%	122	81.3%	127	84.7%
False Positives	22	64.7%	23	67.6%	42	77.8%	28	70.0%	23	67.6%
False Negatives	5	3.8%	6	4.5%	5	4.4%	5	3.9%	6	4.5%
Number of observations	167		167		167		167		167	

** Classifications by the IAP in 1993.*

*** A model was estimated on 383 projects submitted to the IAP between 1989 and 1992 and tested for prediction accuracy on 167 projects submitted to the IAP in 1993.*

able to correctly classify, ex ante, both successful and unsuccessful ventures from the perspective of whether they will reach the market or not. This is an unexpected result as it implies that the experts are able to detect and appropriately use information in a highly multivariate setting where there is a low signal-to-noise ratio and feedback on their decisions is not readily available. Nevertheless, a statistical model with four criteria is able to achieve the same hit-rate as the experts.

One should be clear about the limitations. The statistical model applies to what it has been calibrated on: screening of seed and early-stage investments. It does not apply to other investments, it does not measure return on investment and it does not substitute for due diligence investment review once the initial screening has been undertaken. Nevertheless, it does provide a great improvement over current practice in the screening of seed and early-stage investments.

With some training it is possible for fairly novice assessors at VC firms to use such a model. It is even possible that entrepreneurs themselves provide the information over Web-based forms to further streamline the screening process. For example, such screening processes are currently in place at Garage.com, Canadian Science and Technology Growth Fund and Launchworks Inc., although the latter two funds do not use a statistical model to weight together the evidence. If the IAP's hit-rate is maintained by VCs using my model, it would suggest a doubling of the rate of return on VCs' investments on seed and early-stage investments due to the improvement in the screening ability.

It is plausible that superior statistical decision-support models can be constructed on historical data from other types of investments such as second and third round financing. It's also possible to focus on specific industries. It is obvious that the key criteria may shift across the types of investments. A statistical decision-support model for a non-early-stage fund is presently being developed. Self-selection is a statistical problem that needs to be addressed. Those investments that received funding are more likely to succeed than those that did not, due to the capital injection. However, there are appropriate statistical models and methods to control for this effect. ■