

## DEALING WITH HETEROGENEITY IN ENTREPRENEURSHIP RESEARCH

*Per Davidsson: Queensland University Of Technology, Brisbane, Australia*

~

**Contact:** Per Davidsson, Queensland University of Technology, Gardens Point Campus, 4001 Brisbane, Australia, (T) +617 3864 2051, (F) +617 3864 1299, Email: per.davidsson@qut.edu.au

### ABSTRACT

It has been argued that entrepreneurship research has not “come far enough, fast enough” (Low, 2001). Although the volume of research has grown dramatically, as has the number of specific sub-topics covered, there are relatively few issues on which the entrepreneurship research community has arrived at a consensus. Arguably, one factor that hampers knowledge development in entrepreneurship is the heterogeneity of the phenomenon. Business ventures are started by individuals and teams with very different backgrounds and motivations, pursuing business ideas of very variable inherent quality in environments that also show tremendous variability. As a result it is difficult to arrive at broadly valid generalizations, and studies that include all the variance at once risk arriving at nothing but weak results and increased bewilderment. This paper discusses strategies for dealing with the entrepreneurship phenomenon’s inherent heterogeneity in research design, sampling, operationalization and analysis, so that individual studies can contribute to knowledge accumulation rather than to increased confusion. Specific instances of exemplary, published works that have employed these strategies are discussed throughout the paper.

### A FORMAL LOOK AT THE HETEROGENEITY PROBLEM

Just like any other research, most entrepreneurship research deals at least in part with how one or more circumstances or factors (“explanatory variables”) contribute to produce one or more outcomes (“dependent variables”). The centrality and explicitness of focus on such causal relationships vary across research approaches and paradigms, but some ambition to suggest, unveil, or understand how various factors relate to entrepreneurial outcomes is usually represented in the research. For example, we may want to understand why individuals engage in a business start-up; what makes them persist in entrepreneurship through come-back after failure or becoming a habitual entrepreneur; or what makes them successful in entrepreneurial endeavours. Alternatively, we may be after explanations for firms’ differential growth and seek these in the resources, business model, and other characteristics of the firm itself as well as the conditions of its industry and its regional environment. In other cases still we may want to understand how the institutional conditions influence levels and content of entrepreneurial activity in a country.

While possibly embracing also other aspects all of these research interests include an element of causal relationships as displayed in Figure 1. As drawn, this figure depicts the simplest possible case (barring cases with fewer explanatory variables) where a number of explanatory variables ( $x$ ) have direct and additive effects on the dependent variable ( $y$ ). There is no indication that the explanatory variables are inter-correlated with one another. Hence, as a starting assumption (to be relaxed) we may assume they are truly “independent” variables. Dotted circles and arrows represent variables that are not included in the study, while the thickness of an arrow reflects the true, relative strength of its influence.

The fact that as many as  $n$  factors influence  $Y$  makes the research problem complex, especially as we may not know a priori what all those  $n$  factors are. This may lead us to try a qualitative approach: in order to obtain a complete understanding of the factors that influence  $Y$  in a particular case we need to take both a broad and a deep look at that case. However, here we quickly run into trouble. Even in the unrealistically simple situation depicted in the figure, when the number of variables considered is expanded to  $n$  the number of cases has to be greater than  $n$  for it to be even theoretically possible to tease out the absolute and relative influence of each variable on  $Y$ . In reality the problem will, of course, be much worse. First, the explanatory variables will be correlated and perhaps even causally inter-related, exponentially increasing the mathematical impossibility of solving the problem. Second, in

qualitative research the variables are typically not formally measured, so there is no “hard” way to estimate the relationships anyway. Third, the extent to which the studied cases are representative of the theoretical category or relevant empirical population is typically unknown. In short, qualitative research is not well adapted for assessing the absolute or relative influence a set explanatory variables has on one or more dependent variables. When researchers go that route they engage in intuitive calculations of complex relationships based on informally assessed variables as well as possible over-extension of the generalisability of what they think they have found. Qualitative research can be good for other things, like finding out what explanatory variables are worthy of consideration at all, and suggesting by what mechanisms they might influence the outcome—or finding altogether new research questions worthy of further investigation. Obtaining valid, general estimates of strengths of influence such studies can not.

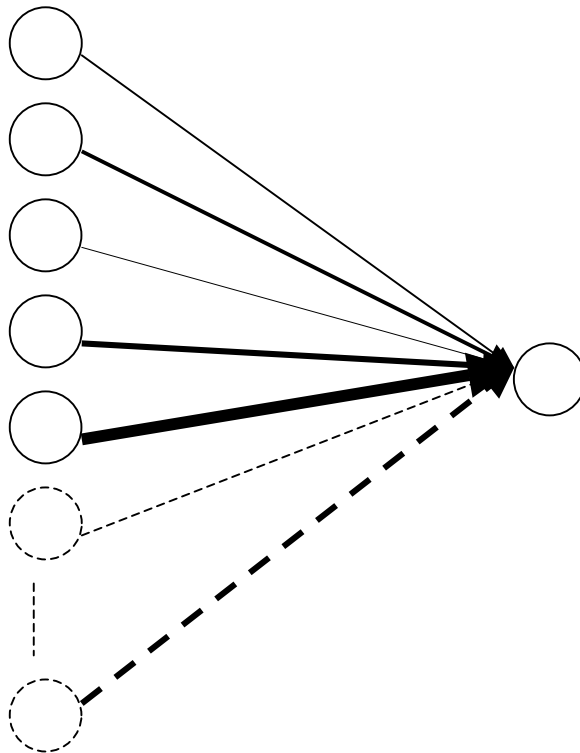


Figure 1. A representation of a multivariate causal relationship

Unfortunately, the heterogeneity problem is not automatically solved merely by turning to a quantitative approach. Very far from that, actually. First, the interrelationships among the variables will not be as simple as in Figure 1. Variables  $X_1$ - $X_6$  will typically be interrelated with one another. This is the problem multivariate statistical analysis methods have been designed to solve. Although they do so relatively well there is some amount of arbitrariness regarding how explanatory power is distributed among the variables. Hence, we may not come out with the right “thickness of the arrows” even in this simple case. Second, variables  $X_7$ - $X_n$  are also likely to have some influence on  $Y$ . Excluding them means reducing explanatory power. This is a problem if full explanation is the goal but not much of a problem if the goal is to test a particular influence suggested by a particular theory.

Third, variables  $X_7$ - $X_n$  are also likely to be correlated with  $X_1$ - $X_6$ . If so, we have a confounding variables problem (Kish, 1987). This means the estimated effects of  $X_1$ - $X_6$  will be distorted. To solve that problem we may want to turn confounding variables into (included) control variables instead. This increase in the number of variables leads to loss of degrees of freedom and demands for increased sample size. While that is a solvable problem there is a worse, accompanying problem that may not be possible to solve. In practice there is a trade-off in quantitative research between the number of variables you can include on the one hand, and how well you can measure them on the other. This is what Figure 2 is getting at.

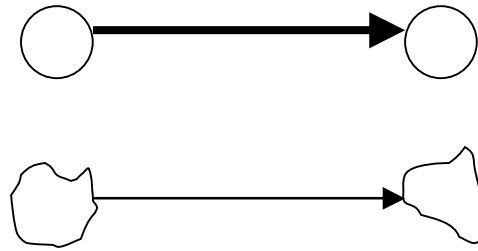


Figure 2 A graphical representation of measurement quality and estimated relationship strength

The upper panel depicts the true, strong relationship that would be estimated if both variables had perfect operationalisations. The lower panel depicts the weakened relationship that will be estimated in the face of measurement error. Consequently, strong relationships may appear weak and weak relationships may appear non-existent. As a result explanatory power will be reduced. If the measurement quality problem hits different variables unequally we will also get an incorrect assessment of relative importance. This is particularly likely when comparing, e.g., the influence of relatively easy-to-measure socio-demographic variables with that of shorthand versions of harder-to-assess psychological constructs.

Fourth,  $X_1$ - $X_n$  are likely to be not only correlated, but causally inter-related. This is the problem that structural equations modelling (SEM) and related techniques have been developed for (Fornell & Larcker, 1981). These are great, sophisticated tools but they can only do so much as regards telling exactly what causal structure is “truer.” Strong theory is needed in order to know which models are worth examining at all and even so the techniques give only indications rather than definitive answers concerning the correctness of the model.

Finally, even if we could estimate it correctly, the effect of a variable  $X_i$  may not be uniform across cases. This is a fundamental principle of Economic Psychology (or Behavioural Economics): the effect of an economic stimulus is contingent on the motivation and ability of the agent, both of which can be measured (Katona, 1975). From psychology proper we recognise this argument as the case for stimulus-organism-response (SOM) models rather than the simpler stimulus-response (SR) models. What this means is that, e.g., the regression coefficients obtained in a multiple regression analysis represent *average* effects for the studied population. Around that average there will be some variation, leaving variance in the dependent variable unexplained even if all relevant explanatory variables have been included and perfectly measured. For example, the level and content of the “human capital” of the founders(s) may be much more important for high-tech start-ups than for new franchise outlets, and the owner’s growth aspirations may have differential effect on real growth when we compare expanding and stagnant industries. If the heterogeneity problem is such that there is huge variability not only in the values of the explanatory variables but also in what variables matter and how much they matter we are likely to end up with an explanatory model with a long list of variables, all of which seem to have some influence—given the sample is large enough to ensure statistical power, (Cohen, 1988)—but none having a very dominant role. Qualitative research approaches are fundamentally more geared towards accepting such variability in what factors matter but do not help us much as regards how to generalise such insights across heterogeneous populations.

Thus, it seems to be the case that no matter what approach we take, heterogeneity is going to haunt us and make the interpretation of our research findings questionable. Unless, that is, we make deliberate efforts to reduce or otherwise account for that heterogeneity. The following sections will discuss how some of the above problems related to heterogeneity can be, and successfully has been, addressed in entrepreneurship research.

### **REDUCING HETEROGENEITY THROUGH BASIC DESIGN: LABORATORY RESEARCH AND NATURAL EXPERIMENTS**

Reducing heterogeneity lies at the heart of experimental research. In an experiment the researcher creates a situation where the variability in the explanatory variable(s) of interest can be manipulated by giving some subjects a “treatment” and others no treatment a different treatment (or a different treatment). The risk of confounding the experimental variable with other influences is controlled by random assignment or matching of subjects to treatments. As an extra precaution, critical

control variables are measured and included as covariates in the analysis. In practice, heterogeneity is often further reduced by using subjects from a homogenous socio-demographic group (like undergraduate students). Other laboratory research like simulations shares some of these heterogeneity-reducing qualities.

Therefore, such approaches would appear to present attractive opportunities for entrepreneurship research. Yet, until recently very little entrepreneurship research used experimental approaches (Chandler & Lyon, 2001). A valid reason for this, of course, is that many research problems do not lend themselves to experimentation for practical or ethical reasons. For example, we cannot get governments to systematically vary the institutional framework across regions according to an experimental design; make people and businesses stay in the regions they have been randomly assigned to, and expect them to behave as if it were a uniform, permanent institutional change rather than an experiment. However, with some creativity quite a range of entrepreneurship issues *can* be addressed through experiments and other methods using laboratory control. Robert Baron has outlined a range of issues related to entrepreneurial knowledge and opportunity recognition that can be addressed with psychological theory and experimental method (Baron, 2006; Baron & Ward, 2004) and Dean Shepherd has championed empirical work in this and other areas using conjoint analysis and other experimental approaches (Shepherd & DeTienne, 2005; Shepherd & Zacharakis, 1997). Laboratory work also underlies Sarasvathy's development of Effectuation Theory (Sarasvathy, 1999, 2001). In line with Baron's (2006) suggestion, Gustafsson (2004) addressed the promising avenue for experimental research of comparing expert and novice entrepreneurs. Many other examples have also been presented at recent conferences but not yet reached other forms of publication. In common they have the quality that they produce relatively clear answers that are likely to be replicable.

When a laboratory approach is not feasible there may be natural settings that share some of the heterogeneity-reducing qualities of laboratory approaches. This might entail external shocks that are equal to all; island economies, and the like. Shane's (2000) study, which some of the previously mentioned experimental research builds on, is one example. That study included all ventures (and their founders) associated with one and the same basic technological innovation. By keeping the basic innovation constant and including the entire "population" the risk that unmeasured heterogeneity distorts the results could be minimised.

A study that has some "island economy" qualities is Usher and Evans' (1996) investigation of the transformation of the petrol station population in Calgary. This narrow industry and region focus eliminates much of the heterogeneity problem and therefore allows a stronger test of the theoretical issues at hand. A somewhat similar example—regrettably only partly available in English—is Karl Gratzer's in-depth reconstruction of the emergence, life, and demise of the Automated Restaurant industry in Sweden (Gratzer, 1999). Another empirical context I have had described to me and which would allow a similar study is a valley in Italy, which used to be populated by almost identical, small, family-owned wineries, all producing the same type of wine. Over a number of decades, some have become acquired or remained the same while others have grown organically and others still have expanded through related or even unrelated diversification. The controlled context and common origin present a reasonable potential for historical case studies to tease out the relative importance factors at different levels of analysis in determining these different development trajectories. With a more diverse set of cases it would be an impossible task. Finally, the usual comparisons of "entrepreneurs" (business owners or founders) with "others" typically confound factors that make people *engage* in entrepreneurial endeavours with those factors that make them *persist* and *succeed* at such tasks, respectively (Davidsson, 2004: 70, 2006a). I recently come across on-going research on what entrepreneurial firms do and do not re-open in New Orleans after the Katrina disaster. Tragic as it is, the post-Katrina situation presents a cleaner context for addressing the specific issue of entrepreneurial persistence.

In short, experimental and experiment-like situations offer conditions of radically reduced heterogeneity, which allows obtaining a clearer image of the key relationships investigated. By moving from laboratory experiments to laboratory-like field situations we have already drifted towards the issue to be discussed next: sampling strategies for reduced heterogeneity.

## DEALING WITH HETEROGENEITY IN SAMPLING

When a field of research is young and its focal phenomenon relatively unexplored, the researcher's typical inclination is to want to study the phenomenon in "its entirety" and to obtain as "complete" an understanding of it as possible. This typically gears the researcher towards a census study or to using a random sample from the entire, empirical population. That sample is likely going to be very heterogeneous; i.e., include all types of start-ups from home-based, part-time efforts of making

some extra income with a “me-too” firm to high-powered, high-tech and high-risk ventures that can potentially change the world—and everything in-between.

As a field matures and becomes more theory driven, researchers typically give up aspirations to completeness. The interest is focused on whether or not a particular factor, suggested by a particular theory, has an important effect—not on complete explanation of the phenomenon. The researcher may also realise that trying to obtain a representative sample from the entire, relevant population is a futile undertaking because that population simply does not exist in one place at one time (Davidsson, 2004: 68-69). We expect the theory to be valid last year and next year and in other places, too. Hence, the task is to find *a* relevant sample to test the theory on. As the researcher is only interested in one particular factor or a limited set of variables, other factors should best be kept constant or be unrelated to the focal relationship, like in experimentation. The solution is to work with a very narrow sample: only one industry, and/or age group, and/or region. This follows the logic that if the theory is any good, it should make correct predictions in this narrow sample. It should also make correct predictions in a range of other contexts not investigated, but that is issue for replication, and possibly for analytical rather than statistical generalisation.

Recent examples show that theory-driven research on narrow sample can lead to strong results regarding relationships that have appeared weak or inconsistent in previous research. An exemplary study in that regard is Baum and Locke (2004). In their study of how individual level psychological variables influence the growth of young firms they use a sample not from the entire small business population and not from, e.g., all of manufacturing, but a much narrower category: North American architectural woodwork firms. There are several reasons for the relative success of this study: possibly a better selection of psychological “traits” than some other studies have used; better than usual operationalisations of the key constructs, and using a structural equations modelling technique for analysis (applying the more reasonable assumption that general, “distal” traits would have indirect rather than direct effects) and a longitudinal design that gave the explanatory variables a reasonable period to show their effect. However, designing away potentially blurring heterogeneity by focusing on a narrow industry was no doubt an important contributor.

In fact, entrepreneurship research that achieves publication in high-prestige disciplinary or mainstream outlets is often of this kind. The studies by Usher & Martin (1996) and Shane (2000) have already been mentioned. Other examples include Eisenhardt and Schoonhoven (1990; semiconductors) and Stuart, Hoang, and Hybels (1999; biotech). In none of these cases is the research driven by a particular interest in these industries. The industries are chosen because they present *a* relevant context for testing theoretical ideas, and to do so without blurring issues by including variation along too many dimensions at once.

Another interesting recent example is Cliff, Devereaux-Jennings and Greenwood (2006). They address the age-old question as to whether it is industry insiders or outsiders who contribute most to innovation in an industry (Schumpeter, 1934). Their results did not support that those with more outside experience were more innovative. However, they did support that those in the periphery rather than the core of the industry were more innovative. Hence, these authors concluded that extensive, core industry experience can indeed constrain innovation. They reached this conclusion based on studying a sample from all law firms created in the Greater Vancouver Regional District from 1990-1998, i.e., and empirical context that is excellently restricted in space, time and as regards type of industry. They also focused their study on a particular type of innovation, namely novelty in organisational design. Like Baum & Locke’s study, Cliff et al.’s research stands out as exemplary also in other respects like careful operationalisations and considering alternative explanations.

However, Cliff et al.’s specific choices also highlight the limitations of studying a narrow empirical context. Baum & Locke’s result for sure require replication before we can accept them as “definitive,” but in their case it is difficult to see why this type of variables would have completely different effects in other industries, countries or periods. But are Cliff et al.’s law firms good representatives for the theoretical category “firms” or “young firms”? The authors themselves see them as representative for highly institutionalised, mature industries and found their selection suitable also because it was easy to identify “core” and “periphery” of the industry. The relative importance of outsiders vs. insiders in may well vary with industry maturity (and with type of innovation, as the authors admit when discussing the limitations of their research).

When the research concerns, for example, the effectiveness of alternative strategies these concerns are aggravated. As the true goal of academic research is not to generalise only to very narrow empirical context, a singular focus on but a particular industry niche in a particular country at a particular time may be less suitable when it is suspected a priori that the influence of the explanatory variables differs by, e.g., industry, firm age, or firm size. However, the answer is not to turn to a simple random sample. First, a simple random sample may be totally dominated by, e.g., tiny, imitative

businesses in mature industries. The mere fact that they are more numerous in a given country at a given time does not make them more important (Davidsson, 2004: 69). Second, simple random sampling does not actively reduce heterogeneity. A better solution may then be to use a stratified sample so that satisfactory representation of different types of firm is ensured while heterogeneity as represented by the idiosyncrasies of non-selected categories is excluded by design.

One example of this approach is Wiklund, Davidsson and Delmar's (2003) study of how expected consequences of growth influence overall growth willingness. They used samples obtained in three different periods of time, representing different business cycle conditions. Further, each sample was pre-stratified into three industries (select sub-groups within manufacturing, retailing and services, respectively) and three size classes. In this way, it could be tested whether or not the overall results held up in each sub-sample. Some of their results are displayed in Table 1. The results in this case reveal similarity across sub-groups in that the strongest predictor remains the same in each analysis, and the variance explained of a comparable magnitude. Further, all effects that are statistically significant are positive, as expected.

Likewise, Brown, Davidsson, & Wiklund (2001) pre-stratified their very large sample by industry, size class, and governance structure (independent vs. subsidiary). In both these example the stratification served primarily to show similarity across sub-groups. In other cases the sub-groups may be chosen because theory suggests the results should be different in the different groups.

Table 1. Expected growth consequences and overall growth willingness (Wiklund et al, 2003)

Sub-sample	Manuf. n=571	Service n=340	Retail n=246	5-9 emp n=326	10-19 emp n=479	20-49 emp n=353	Old firms n=771	Young firms n=372
Variable								
Workload	.07*	.08	.00	.07	.08*	-.01	.07*	.07
Work tasks	.04	.06	.02	.13**	.05	-.05	.01	.10*
Employee well-being	.23***	.27***	.23***	.30***	.17***	.29***	.28***	.22**
Personal income	.10**	.10*	.10*	.11*	.06	.13**	.12***	.05
Control	.08*	.10*	.08	.07	.12**	.08	.10**	.04
Independence	.11**	.09*	.14*	.02	.15***	.13**	.10**	.13**
Survival of crises	.09*	.10*	.04	.07	.07	.13**	.09**	.06
Product/service quality	.06	.04	.09	.06	.09*	.00	.03	.11*
Adj. R <sup>2</sup>	.22	.28	.16	.26	.24	.21	.25	.20

Note: Forced entry of explanatory variables is used. Standardised regression coefficients are displayed in the Table. \* = p < .05; \*\* = p < .01; \*\*\* = p < .001. Single-tailed test of significance is applied.

Needless to say, in order for stratification to be really meaningful the total sample needs to be large enough to allow sub-group analysis with enough retained statistical power. However, that is arguably exactly what the field needs: fewer but more comprehensive research studies.

### DEALING WITH HETEROGENEITY IN OPERATIONALISATION

Heterogeneity of the entrepreneurship phenomenon has consequences for operationalisations, too. For example, the best measure of firm size may be the number of vehicles for a taxi company, the number of seats for a restaurant operation, and the quantity of electricity delivered for a power station. However, how are we to compare the firms' growth across these different measures? Sales and number of employees are more generally applicable, but may have other disadvantages (Bolton, 1971; Davidsson & Wiklund, 2000). Baum and Locke (2004) used the generally applicable measures of sales and employment to assess their dependent variable, growth. In that regard, then it would have been possible to expand their study to include other industries. That would come at a certain risk, though, for what is "high" and "low" depends in part on the industry.

Cliff et al. (2005) used a dependent variable that is much more difficult to compare across industries. How does one measure "innovative intensity" in a way that is comparable across industries, firm sizes, and types of innovation? It is an almost insurmountable task, and when the research concerns new, small firms across industries it is clear that neither formal R&D expenditure nor numbers of patents is a very useful measure (Acs & Audretsch, 1990). Within narrow industries (and for certain types of innovations) it may be possible to arrive at strong operationalisations, which is precisely the strength of Cliff et al.'s design. When a "one-size-fits-all" measure has to be used we end

up with the problem described in Figure 2 above. For example, one of the innovation items in the best established measure of firm level entrepreneurship, the *Entrepreneurial Orientation* scale (Rauch, Wiklund, Lumpkin, & Frese, 2006), reads “How many new kinds of products or services has your company introduced over the past 5 years?” How would “equally innovative” manufacturers and service firm answer this question? As a way to partly overcome the problem of general applicability the high end of the response scale has had to be anchored with a weak quantification: “a lot of new products/services.” This makes it highly subjective on top of still being sensitive to industry and firm size, and as regards retailing firms it is questionable whether the item is meaningful at all.

As a consequence, antecedents and effects of firm innovativeness would be underestimated in studies of heterogeneous samples, either because of a compromise that works “so-so” for all is used or because a more specialised measure is employed, which works excellently for parts of the sample and not at all for other parts of it. Similarly, when Harrison, Mason and Girling (2004) investigated financial bootstrapping behaviour using a sample from the software industry, they included items like “commercialising public domain software”; “porting fees to transfer software from one platform to another” and “using public domain development tools” alongside more general indicators. This likely led to better assessment of bootstrapping in that particular context, but also to developing a measure that cannot be applied elsewhere.

In developing measures of entrepreneurial behaviour and employ those to heterogenous samples the less-than-perfect alternatives we are left with seem to be the following (cf. Davidsson, 2004: 113):

1. Develop one operationalisation that is assumed to be good for all ventures/firms. Accept that interesting manifestations of entrepreneurship that clearly apply only to narrow subsets of firms cannot be included in the measure. Also accept as a fact that larger firms and firms in some industries, on average, exercise more entrepreneurship than do smaller firms and firms in certain other industries.
2. Develop one operationalisation for all ventures/firms. However, normalise the score within industry/size class (or other) groups, and use deviation from the own class mean as the measure of entrepreneurship. This would eliminate what can be regarded as a bias against certain categories when approach (1) is applied, but this comes at the cost of assuming that all subgroups of firm are equally “entrepreneurial.” That is, only within-group and not between-group differences will be detected.
3. Develop separate and adapted operationalisations for different sub-groups (by industry, size class, or otherwise). Standardise this measure, so that comparisons can be made across different operationalisations of entrepreneurship. This would allow including the presumably most relevant indicators for each category, but this type of procedure is not well established and involves a considerable risk of comparing apples and oranges in the analysis.

Similar issues could be discussed with regard to variables believed to explain or result from entrepreneurial behaviour. The problem is a very difficult one and all things considered it may well be worth focusing on a narrower empirical context also for reasons of operationalisation.

## **DEALING WITH HETEROGENEITY IN THE ANALYSIS**

For some entrepreneurship research problems it is very difficult to effectively address heterogeneity at the design stage (other than using generally applicable operationalisations). A case in point is research like the *Panel Study of Entrepreneurial Dynamics* (PSED) (Gartner, Shaver, Carter, & Reynolds, 2004). This research attempts to study representative samples of emerging ventures. Due to the emerging nature of these entities, they cannot be pre-stratified by industry, size, or otherwise. Whether or not heterogeneity has been reduced by design, however, there is a number of approaches to dealing with it in analysis. This includes *accepting partial explanation; including control variables*; performing separate, *sub-sample analyses*, and various ways of *modelling heterogeneity*. Each of these will be briefly discussed below.

### **Accepting partial explanation**

In table 1, it is revealed that the eight types of expectations used in the study explain some 20-25 percent of the variance in overall growth willingness. Is this because there are other consequences of growth that are more important than those included but the researchers were not sensible enough to include? Hardly, their selection was based on the best available knowledge at the time, using both an extensive survey of the literature and a pilot study to find the right dimensions. One that arguable should have been included from start (expected value growth of the firms) was added in the last survey

without adding much. Various control variables (other explanations than expectations) have also been tried without substantially improving the results.

Measurement error (remember Figure 2) is likely to have deflated the results, especially as each “expected consequence” was assessed with a single item. However, I would argue that even with perfect of all variables and inclusion of all reasonable control variables, total explanatory power would be unlikely to exceed 50 percent. This is for a reason mentioned already in the front part of this paper: the respondents differ not only as regards the true values on the explanatory and dependent variables, but also as regards the true strength of the relationships between the two. Depending on who they are and their current situations, expectations that growth would lead to an increased workload may be an absolute growth deterrent for one business owner but inconsequential for another. Thus, the coefficients represent average effects, leaving variance unexplained. This is likely to be the case in most social science research, and hence one approach to dealing with heterogeneity in the analysis is to accept its deflationary influence on explanatory power.

### **Including control variables**

One way of dealing with heterogeneity is to include control variables in the analysis. For example, the previously used examples of Baum and Locke (2004), Cliff et al. (2006) and Wiklund et al. (2003) all use a set of 4-7 control variables in their respective full sample analyses, representing key characteristics of the founders (e.g., age, gender, education), the firm (e.g., age, size), and the environment (e.g., munificence). Failing to include potential control variables that co-vary only with the dependent variable has no serious consequences as it merely reduces the amount of total variance explained. The effects ascribed to the included explanatory variables remain the same. It is when the control variables correlate with other explanatory variables that their inclusion is critically important. For example, despite significant zero-order correlations there appear to be few pure “gender effects” but only indirect effects via, e.g., choice of industry or amount of resources invested. When these controls are included the gender effect typically disappears (DuRietz & Henrekson, 2000; Watson, 2002). Similarly, because larger firms simply do more things, the hypothetical variable “bureaucratisation” may be attributed a positive effect on innovativeness if a count based measure of the latter is employed. If a control for firm size (likely to be highly correlated with bureaucratisation) is included this effect will likely be attributed to size instead, and the effect now attributed to bureaucratisation may well be the expected negative one, in spite of a positive zero order correlation.

For the reviewed reasons control variables should be used to reduce the adverse effects of heterogeneity. It is also important to realise, however, what control variables do *not* help us with. In Table 1 there is some indication that expectations concerning the feeling of *independence* are rather inconsequential in the smallest size group whereas it is of some importance among the somewhat larger firms. This would not show in a full sample analysis including a control variable for firm size—we still only get one, average estimate for the influence of independence expectations on overall growth willingness. In the hypothetical case that the effect of a variable were completely different for men and women—say strongly negative for the former and mildly positive for the latter—a full sample analysis with gender as control would (under mild assumptions) give us a weak negative effect for the variable in question (which is true for neither group) and a positive effect for the control variable denoting female sex. That would not be very helpful for arriving at a correct interpretation of the relationships. In order to tease out the true effects by sex the groups would either have to be analysed separately, or the relationship would have to be modelled as moderated by gender. These are possibilities we will turn to in the following two sub-sections.

### **Performing sub-sample analysis**

The merits of performing separate sub-sample analyses were partly covered above while discussing stratified sampling. Even if pre-stratification is not possible the analysis can be broken down by categories in the analysis, provided that the sub-groups are large enough. In Table 1, this is the case for the analysis by firm age in the rightmost columns, as age was not used as a stratification variable at the sampling stage.

The advantages of sub-group analysis should be pretty obvious, provided the analysed sub-groups are large enough. If they are small seemingly large (but statistically insignificant) differences may appear simply because of random sampling variation. For some purposes analysis by subgroups may be critically important. For example, Shane and Venkataraman (2000) observed that whereas entrepreneurship researchers have been preoccupied with assessing variance among the individuals who start businesses there is a dearth of research focusing on the effects of variance in the venture ideas (or “opportunities” as Shane and Venkataraman call them) that they pursue. Potential importance of this is demonstrated by Mikael Samuelsson’s dissertation research (Samuelsson, 2001, 2004). Using

the Swedish PSED counterpart data, this study post-stratified innovative versus imitative ventures. The results revealed that the set of variables associated with making progress in the venture creation process was very different for the two types of ventures. Moreover, while the model was reasonably successful at explaining outcomes for the innovative ventures the amount of variance explained was very modest for the much larger group of imitative start-ups. This was in spite of including a rather broad range of explanatory variables. Although remaining high levels of heterogeneity in the imitative group may be part of the explanation these results are also an invitation to a major rethinking of the factors that lead to successful venture creation for the imitative majority.

### **Modelling heterogeneous effect**

There are different ways to model expectation of differential effects of one explanatory variable depending on the value of another explanatory variable. The arguably most common approach is to include interaction terms in a regression analysis; so called “moderated regression.” The popularity of this approach has increased dramatically recently in entrepreneurship. As a case in point, every issue of the 2006 volume of *Journal of Business Venturing* includes one or more articles applying some form of analysis of interaction effects. This has been rewarded with considerable success; in my perception more so than earlier, similar attempts at examining contingent relationships in research on larger organisations. The reasons for this can only be speculated. Possibly, while complex enough in their own right, small and emerging ventures are less complex than large organisations, making it more manageable to include and assess the most critical contingencies in the study.

Analysis of interaction effects basically solves the same problem as separate sub-group analysis does. In fact, when the interaction includes a categorical variable and the sample is large enough using sub-group analysis is an easier way to reveal and communicate the results. However, when the sample size does not suffice for all required breakdowns; when the interaction is between two continuous variables, or when the analysis involves three-way or higher order interaction terms it will likely be more practical and economical (in terms of power and degrees of freedom) to perform the analysis as a moderated regression analysis.

Rather than choosing an example from published research in a prestigious journal—of which there are many examples—I will in this case use results obtained by Hmieleski and Ensley (2004) in what is best described as a pilot study about the effects of different forms of intelligence. I use this example because it so powerfully illustrates the *type* of revelations that an analysis of interaction effects can yield, and how it can lead to enhanced sense-making. In short, according to their results analytical intelligence has no main effect on venture outcomes in the pre-formation and formation stages. However, this conventional form of intelligence has a strong positive effect on performance in the presumably more structured and less genuinely uncertain growth stage. For creative and practical intelligence, presumably conducive of early market experimentation, the pattern is the opposite—they are ascribed positive effects early on. The really interesting interaction is that the positive effects of these latter types of intelligence in the earliest stages are boosted if analytical intelligence is also high. That is, the results suggest that people high *only* on analytical intelligence are not helped by this in the early, highly uncertain stages of venture development. For those who possess creative and practical intelligence, however, a sound dose of analytical intelligence helps direct these other talents to more productive endeavours. This, to me, is a highly intriguing pattern of relationships that also seems to make a great deal of sense. Bearing in mind that the study has limitations as regards potential retrospection bias and the operationalisation of the hard-to-capture intelligence variables, a replication with strengthened methodology would be needed in order to accept it as established truth.

Eckhardt, Shane and Delmar (2006) represent a recent example of a different and less common approach to modelling heterogeneity. These authors apply multi-stage selection modelling to the problem of predicting which new ventures receive external funding. Conventionally, such a research question would include some characteristics of the founders and hopefully some characteristics of the venture—and possibly some interaction between the two—in the same regression analysis. Logically, however, receiving external funding requires that the founders actively seek such funding. Therefore, Eckhardt et al. (2006) in the first stage hypothesise that variables reflecting founders’ subjective assessment of the future outlook for the venture (perceived market growth; expected employment size, and expected price competition) determine whether external finance will be sought or not—and estimate these relationships. In a second stage they hypothesise (and estimate) that objectively verifiable characteristics of the venture (whether sales have already been obtained and whether various “start-up activities” have been completed) will determine external investors’ willingness to fund the venture, *given* that financing is sought.

The results come out different but in their particular case not very markedly different from a model where external funding is regressed on both founder perceptions and venture characteristics in a

single analysis. However, in principle the analytical approach the Eckhardt et al. (2006) study uses is hugely important because it speaks to the central fact that *entrepreneurship requires human agency* (Shane, 2003). In many cases, other variables can have their effects only if the entrepreneur chooses to let them have their effects. The multi-stage selection approach then appears more valid than the more gradual thinking that underlies modelling of interaction effects. In other cases human decision is not crucial or the entrepreneur's reluctance more negotiable and interaction modelling may then be the more valid approach.

Finally, a somewhat more idiosyncratic but hugely impressive example of modelling heterogeneity is the study by Gimeno, Folta, Cooper and Woo (1997). Their study is about prediction of venture survival, and in attacking this problem they develop solutions to a problem that many other researchers have not dealt with properly, therefore arriving at confusingly weak or even counterintuitive results. This problem is that general human capital, as reflected in general education and management experience, should not only have a positive influence on a venture's economic performance, but also on the attractiveness of other alternatives (such as employment or starting a different venture instead). Therefore, these authors argue that the effect of general human capital on the likelihood of exit will be indeterminate a priori. While having a positive effect on economic performance it will also have a positive effect on the minimum threshold of performance judges sufficient to stay in business. This is an important insight that much other research has neglected (Davidsson, 2006b). However, the most impressive aspect of the study is the way in which Gimeno et al. (1999) manage to overcome the problem of dealing with an unobserved threshold level (and other empirical problems) by combining elements of two known techniques (tobit modelling or censored regression, and grouped data regression) both of which separately would have been relative novelties in entrepreneurship research.

The results show mixed support for the central hypothesis. Management experience does indeed seem to be associated with a higher threshold level, which explains its insignificant relationship with likelihood of exit. Education and supervisory experience appear negatively related to exit largely because of a higher payoff in the venture. A separate test suggested that the expected payoff in alternative employment was higher for managerial than for supervisory experience. If it cannot be demanded that the average future study matches the rare conceptual and methodological sophistication that Gimeno et al. demonstrate it may be reasonable to suggest that greater consideration of (heterogeneity in) the *other alternative* is the minimum that future research should learn from their study.

## CONCLUSION

Due largely to its inherent heterogeneity, different studies of aspects of entrepreneurship do not always add up to a coherent image of the entrepreneurship phenomenon. Not even studies that seemingly address the same aspect of the phenomenon always accumulate to a generally accepted "received view." This paper has discussed various ways to deal with the heterogeneity problem in entrepreneurship research. When piled up on top on one another all the issues discussed may easily portray the image that conducting useful entrepreneurship is an impossible task. I hope that is not the image I have conveyed. My purpose has been constructive. While *perfect* research is not, and never was, within reach for human mortals there are manageable ways in which we can increase the usefulness of our research. I hope the examples highlighted in this paper will inspire others to squeeze more out of their research efforts.

## References

- Acs, Z. J., & Audretsch, D. B. (1990). *Innovation and Small Firms*. Cambridge, MA: MIT Press.
- Baron, R. (2006). Opportunity recognition as pattern recognition: How entrepreneurs "connect the dots" to identify new business opportunities. *Academy of Management Perspectives*, 20(1), 104-119.
- Baron, R., & Ward, T. (2004). Expanding entrepreneurial cognition's toolbox: Potential contributions from the field of cognitive science. *Entrepreneurship Theory and Practice*, 28(6), 553-573.
- Baum, J. R., & Locke, E. A. (2004). The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth. *Journal of Applied Psychology*, 89(4), 587-598.
- Bolton, J. E. (1971). *Small Firms. Report of the Committee of Inquiry on Small Firms*. London: Her Majesty's Stationery Office.

- Brown, T., Davidsson, P., & Wiklund, J. (2001). An operationalization of Stevenson's conceptualization of entrepreneurship as opportunity-based firm behavior. *Strategic Management Journal*, 22(10), 953-968.
- Chandler, G. N., & Lyon, D. W. (2001). Methodological issues in entrepreneurship research: the past decade. *Entrepreneurship Theory & Practice*, 25(4, Summer), 101-113.
- Cliff, J. E., Devereaux-Jennings, P., & Greenwood, R. (2006). New to the game and questioning the rules: The experiences and beliefs of founders who start imitative versus innovative firms. *Journal of Business Venturing*, 21, 633-663.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Davidsson, P. (2004). *Researching Entrepreneurship*. New York: Springer.
- Davidsson, P. (2006a). Method challenges and opportunities in the psychological study of entrepreneurship. In J. R. Baum, M. Frese & R. A. Baron (Eds.), *The Psychology of Entrepreneurship* (pp. 287-323). Mahway, NJ.: Erlbaum.
- Davidsson, P. (2006b). Nascent entrepreneurship: Empirical studies and developments. *Foundations and Trends in Entrepreneurship*, 2(1), 1-76.
- Davidsson, P., & Wiklund, J. (2000). Conceptual and empirical challenges in the study of firm growth. In D. Sexton & H. Landström (Eds.), *The Blackwell Handbook of Entrepreneurship* (pp. 26-44). Oxford, MA: Blackwell Business.
- DuRietz, A., & Henrekson, M. (2000). Testing the female underperformance hypothesis. *Small Business Economics*, 14(1), 1-10.
- Eckhardt, J., Shane, S., & Delmar, F. (2006). Multistage selection and the financing of new ventures. *Management Science*, 52(2), 220-232.
- Eisenhardt, K. M., & Schoonhoven, C. B. (1990). Organizational growth: Linking founding team, strategy, environment and growth among US semiconductor ventures, 1978-1988. *Administrative Science Quarterly*, 35(3), 504-530.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39-50.
- Gartner, W. B., Shaver, K. G., Carter, N. M., & Reynolds, P. D. (2004). *Handbook of Entrepreneurial Dynamics: The Process of Business Creation*. Thousand Oaks, CA: Sage.
- Gimeno, J., Folta, T. B., Cooper, A. C., & Woo, C. Y. (1997). Survival of the fittest? Entrepreneurial human capital and the persistence of underperforming firms. *Administrative Science Quarterly*, 42, 750-783.
- Gratzer, K. (1999). The making of a new industry -- the introduction of fast food in Sweden. In B. Johannisson & H. Landström (Eds.), *Images of Entrepreneurship Research -- Emergent Swedish Contributions to Academic Research* (pp. 82-114). Lund, Sweden: Studentlitteratur.
- Gustafsson, V. (2004). *Entrepreneurial Decision-Making*. Doctoral dissertation. Jönköping: Jönköping International Business School.
- Harrison, R. T., Mason, C. M., & Girling, P. (2004). Financial bootstrapping and venture development in the software industry. *Entrepreneurship and Regional Development*, 16, 307-333.
- Hmieleski, K. M., & Ensley, M. D. (2004). *An investigation of the linkage between entrepreneur intelligence and new venture performance*. Paper presented at the Babson College/Kauffman Foundation Entrepreneurship Research Conference, Wellesley, MA.
- Katona, G. (1975). *Psychological Economics*. New York: Elsevier.
- Kish, L. (1987). *Statistical design for research*. New York: Wiley.
- Low, M. (2001). The adolescence of entrepreneurship research: specification of purpose. *Entrepreneurship Theory & Practice*, 25(4, Summer), 17-25.
- Rauch, a., Wiklund, J., Lumpkin, t., & Frese, M. (2006). *Entrepreneurial orientation and business performance: A meta-analysis*. Working Paper. Giessen: Justus-Liebig University.
- Samuelsson, M. (2001). Modelling the nascent venture opportunity exploitation process across time. In W. D. Bygrave, E. Autio, C. G. Brush, P. Davidsson, P. G. Greene, P. D. Reynolds & H. J. Sapienza (Eds.), *Frontiers of Entrepreneurship Research 2001* (pp. 66-79). Wellesley, MA.
- Samuelsson, M. (2004). *Creating New Ventures: A Longitudinal Investigation of the Nascent Venturing Process*. Doctoral dissertation. Jönköping: Jönköping International Business School.
- Sarasvathy, S. (1999). Decision making in the absence of markets: An empirically grounded model of entrepreneurial expertise. School of Business: University of Washington.
- Sarasvathy, S. (2001). Causation and effectuation: towards a theoretical shift from economic inevitability to entrepreneurial contingency. *Academy of Management Review*, 26(2), 243-288.
- Schumpeter, J. A. (1934). *The Theory of Economic Development*. Cambridge, MA: Harvard University Press.

- Shane, S. (2000). Prior knowledge and the discovery of entrepreneurial opportunities. *Organization Science*, 11(4), 448-469.
- Shane, S. (2003). *A General Theory of Entrepreneurship: The Individual-Opportunity Nexus*. Cheltenham, UK: Edward Elgar.
- Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review*, 25(1), 217-226.
- Shepherd, D., & DeTienne, D. (2005). Prior knowledge, potential financial reward, and opportunity identification. *Entrepreneurship Theory and Practice*, 29(1), 91-112.
- Shepherd, D., & Zacharakis, A. (1997). Conjoint Analysis: A window of opportunity for entrepreneurship research. In J. Katz & R. H. Brockhaus (Eds.), *Advances in Entrepreneurship, Firm Emergence, and Growth* (Vol. 3, pp. 203-248). Greenwich, CT: JAI Press.
- Stuart, T. E., Hoang, H., & Hybels, R. C. (1999). Interorganizational endorsements and the performance of entrepreneurial ventures. *Administrative Science Quarterly*, 44, 315-349.
- Usher, J. M., & Evans, M. G. (1996). Life and death along gasoline alley: Darwinian and Lamarckian processes in a differentiating population. *Academy of Management Journal*, 39(5), 1428-1466.
- Watson, J. (2002). Comparing the performance of male- and female-controlled business: Relating outputs to inputs. *Entrepreneurship: Theory & Practice*, 26(3), 91-100.
- Wiklund, J., Davidsson, P., & Delmar, F. (2003). What do they think and feel about growth? An expectancy-value approach to small business managers' attitudes towards growth. *Entrepreneurship Theory & Practice*, 27(3), 247-269.