DEALING WITH HETEROGENEITY PROBLEMS AND CAUSAL EFFECT ESTIMATION IN ENTREPRENEURSHIP RESEARCH

ABSTRACT

This paper deals with causal effect estimation strategies in highly heterogeneous empirical settings such as entrepreneurship. We argue that the clearer used of modern tools developed to deal with the estimation of causal effects in combination with our analysis of different sources of heterogeneity in entrepreneurship can lead to entrepreneurship with higher internal validity. We specifically lend support from the counterfactual logic and modern research of estimation strategies for causal effect estimation.

1. EXECUTIVE SUMMARY

Entrepreneurship comes in many shapes and forms, driven by a broad variety of motivations and in a diversity of contexts. While this heterogeneity contributes to making entrepreneurship fascinating it also makes it very challenging for entrepreneurship researchers to arrive at strong and credible conclusions regarding causal relationships. Building on experiences from within and outside of entrepreneurship research this article provides an integrated discussion of strategies for dealing with the problem of heterogeneity with particular application to the entrepreneurship domain and the estimation of causal effects. Specifically, we deal with three problems: 1) *unobserved heterogeneity*, i.e., that unmeasured or unavailable variables may bias estimated relationships; 2) *causal heterogeneity,* i.e., that the structure, strength, direction or form of relationships may vary among sub-groups of the studied population, and 3) *uneven validity,* i.e., that the validity of chosen operationalizations may vary by sub-group or context. We discuss how these problems can be reduced at different stages of the research process, i.e., through *theory and theorizing;* in choosing a basic *design* for the study (including sampling)*;* at the *operationalization* stage, and through approaches chosen for *analysis,* respectively. We conclude each section with summarized advice that should help entrepreneurship researchers design more robust studies and arrive at more valid conclusions from extant data sets. Throughout, we illustrate with examples from entrepreneurship studies.

2. INTRODUCTION

This paper deals with causal effect estimation strategies in highly heterogeneous empirical settings such as entrepreneurship. Business ventures are started by individuals and teams with very different backgrounds and motivations, pursuing different objectives based on business ideas of very variable inherent quality in environments that also show tremendous variability. Certain aspects of this great variability or heterogeneity is an important, fundamental and theoretically interesting characteristic of the entrepreneurship phenomenon (Alvarez & Busenitz, 2001; Davidsson, 2004). After all, it is in great part their ability to deviate from norms in new and unexpected ways that makes new and growing ventures fascinating and – sometimes – financially successful. However, the great variability also makes it difficult for researchers to arrive at valid causal inference, and studies that try to 'reflect reality' by including all the multi-dimensional variance at once risk arriving at weak or confusing results. Similarly, studies seemingly addressing the same questions using different samples, operationalizations or analysis approaches may arrive at conflicting results. Over the years, this has led to frustration that "entrepreneurs seem to defy aggregation" (Low & MacMillan, 1988) and that we are "getting more pieces of the puzzle, but no picture is emerging" (Koppl & Minniti, 2003

In this article we use *heterogeneity* as an umbrella term for the simultaneous variability along three different dimensions that makes it challenging to adequately measure theoretical constructs and to correctly model and to estimate causal relationships. Numerous factors contribute to problematic heterogeneity. They are (1) unobserved heterogeneity, (2) causal heterogeneity and (3) uneven or differential validity.

While all scientific fields have to deal to some extent with heterogeneity problems there are several reasons why they are of particular significance for entrepreneurship research. First, the phenomenon itself may be more heterogeneous as it concerns emerging ventures, industries and populations. In more mature stages of development, market forces (Lawless & Tegarden, 1991), learning (Jovanovic, 1982) and institutional pressures (Henrekson, 2007) tend to limit the range of variation along at least some dimensions. Second, due to the cost and difficulty of obtaining primary data on such emerging phenomena researchers may turn to archival data that do not include all variables needed to avoid severe omitted variables problems (Shane, 2006). Third, within the multi-disciplinary field of entrepreneurship research, each theory or discipline emphasizes its specific set of variables and neglects others (Acs & Audretsch, 2003; Ireland & Webb, 2007). Partial absorption of what can be considered unified bodies of theory (and related empirical works) across a range of fields can lead to seriously misreading the theory and results. Many results may only be valid under certain theoretical assumptions and in specific empirical contexts. Hence, the construction of an integrated field of knowledge for entrepreneurship – important as it is – is associated with considerable risks[1].

This article provides an integrated account of the problem of heterogeneity related to causal effect estimation strategies as it presents itself through the entire research process and in an entrepreneurship context. We offer advice based on experiences from within and outside of entrepreneurship research. We also provide an entry point for entrepreneurship researchers to more specialized texts that cover in greater depth particular aspects of the heterogeneity problem from particular disciplinary perspectives but without integration or application to entrepreneurship (e.g., Hausman & Taylor, 1981; Rosenbaum, 2005; Shugan, 2006). We believe such a contribution is timely because evidence suggests that at the present time even the ‘high end’ of entrepreneurship scholarship struggles with heterogeneity problems and

[1] This is a paradox that entrepreneurship as field has to deal with (Ireland, Webb, & Coombs, 2005; Sorensen, & Stuart, 2008). As the field has grown it has moved towards fragmentation and research strongly grounded in theoretical perspectives has become more published (Ireland & Webb, 2007). This is good for the field and new and better knowledge is without doubt getting produced. However, it also makes it more difficult to integrate this knowledge because there is relative little academic premium to doing so compared to producing an empirical paper grounded in a single theory.

causal inference or the identification problem (Shane, 2006). In addition, we provide methods experts with a sense of the typical heterogeneity problems of entrepreneurship research, thereby facilitating their making additional contributions to increased sophistication of this field.

We organize our discussion as follows. In the next section we outline the various types of problems heterogeneity can cause and why this create a specific challenge to estimate causal effects. We then discuss remedies to heterogeneity problems through the use of theory and theorizing. We continue by addressing heterogeneity considerations at the design stage, including sampling strategies. After a discussion of how heterogeneity relates to operationalization we devote the second half of the paper to various remedies that can be applied when analyzing the data. Although we do not quantitatively review the occurrence of heterogeneity problems and specific remedies thereof in published entrepreneurship research we will throughout the manuscript make use of findings from previous method reviews. We will also provide illustrations of how selected entrepreneurship studies have successfully dealt with the heterogeneity issues we raise.

3. HETEROGENIETY AND CAUSAL EFFECT ESTIMATION

Most social research studies, whether quantitative or qualitative, deal explicitly or at least implicitly with causal relationships (King, Keohane, & Verba, 1994). Entrepreneurship is no exception, as illustrated, for example, by Shane and Venkataraman’s (2000: 218) three fundamental questions for entrepreneurship research. Entrepreneurship researchers takes an interest in understanding what personal characteristics make individuals engage, persist or succeed in business start-up activities (Davidsson & Honig, 2003). Alternatively, they seek explanations for differential levels of innovativeness in the characteristics of the firm itself (Cliff, Devereaux-Jennings, & Greenwood, 2006) as well as in the conditions of its regional environment (Maillat, 1998). In other cases still they may want to understand how national institutional conditions influence levels and contents of entrepreneurial activity across countries (Henrekson, 2007).

These examples concern how one or more circumstances or factors (‘explanatory variables’) cause one or more outcomes (‘dependent variables’). In order to illustrate the type of heterogeneity problems encountered in such research we can start with the simple model displayed in Figure 1. In this figure, X denotes theoretical constructs whose influence is to be assessed on the dependent variable(s), i.e., theoretical construct Y. The operationalizations of these constructs are denoted by *X’* and *Y*’, respectively. The solid arrow from X to Y represents the true relationship between the theoretical entities whereas the dotted arrow between *X’* and *Y’* represents the estimates obtained in the research (cf. Bacharach, 1989). We may think of X as the variables whose causal influence on Y we have a theoretical interest in. However, in order to avoid biasing influence of heterogeneity we may also want to include other variables in *X’*. Z represents variables that are not of theoretical interest as disregarding them may lead to a biased picture of X → Y relationships. Some Z variables may be measurable/available and thus possible to include in *X’* whereas other Z variables may be genuinely unobservable.

++++++++++++++++++++++++++

Insert Figure 1 about here!

++++++++++++++++++++++++++

In the situation illustrated by Figure 1 an array of problems may lead to weaker or less correct explanation of Y than expected. Some of these problems are a) low explained variance in *Y’* because of the exclusion of Z; b) low explained variance in *Y’* because (uniformly across the studied population) causation is probabilistic (i.e., imperfectly regular; cf. Yang, 2006); c) low explained variance in *Y’* and under estimated influence of X because of uniformly low validity and/or reliability (large measurement error) of *X'* or *Y'* or both, or d) unreliable estimates of the influence of individual X variables due to high inter-correlations among X. However, neither of these problems are instances of systematic bias due to heterogeneity. In the following, we limit the discussion to problems of the latter kind, namely:

#1. *The problem of unobserved heterogeneity* (e.g., Shugan, 2006). Also discussed under labels such as *omitted variable bias* (Lee, 1982) or *confounding variables* (Kish, 1987), this is the central heterogeneity problem that if Z has substantial correlations with both *X'* and *Y'*, excluding Z will lead to serious bias in the coefficients. The problem is aggravated if the omitted Z variables are causally related with X as the model is then also structurally misspecified (see. #2).

#2. *The problem of causal heterogeneity* (e.g., Western, 1998). Related to the notion of *boundary conditions* (Bacharach, 1989) and with mediation and moderation modeling as solutions in special cases (Baron & Kenny, 1986), this concept denotes the more general problem that the effect of X on Y may not be uniform across elements or subgroups of the studied population. For example, the effect of one X variable may be different in different subgroups, contingent on the value of another continuous X variable, or affect Y indirectly via another X variable.

#3. *The problem of uneven validity. uneven validity*Different aspects of this problem are highlighted particularly in cross-cultural research under labels such as *construct equivalence, instrument equivalence, measurement equivalence* and *measurement invariance* (e.g., Byrne & Watkins, 2003; Schaffer & Riordan, 2003; Singh, 1995). If the X ←→ *X'* and/or Y ←→ *Y'* correspondence – i.e., validity – is uneven it means the chosen constructs and/or operationalizations are not equally suitable to all subgroups of the population. As a result *X'* → *Y'* relationships will be misestimated. That is, this method artifact may be misinterpreted as substantive differences in the nature and strength of relationships.

The omnipresence of these heterogeneity problems may make it impossible to conduct a flawless study of entrepreneurship. However, a researcher who does not try hard to approach that ideal is unlikely to come up with strong and valid findings about the phenomenon. In the remainder of this manuscript we will discuss how these sources of heterogeneity can be dealt with through decisions related to theory, design, operationalization, and analysis.

4. DEALING WITH HETEROGENIET THROUGH THEORY AND THEORIZING

Theory is usually defined as "constructs linked together by propositions that have an underlying, coherent logic and related assumptions" (Davis, Eisenhardt, & Bingham, 2007: 481; cf. Buchanan, 1989: 496, 498) or some variation on that theme. In this section we will show how some well known frameworks regarding theory can be applied in order to address heterogeneity problems in entrepreneurship research. First, we approach the level of maturity in our field as it has an effect on not only how models are constructed and understood, but also how we can handle heterogeneity problems. Second, we explain the functioning of the increasingly popular counterfactual logic. Third, we discuss the specific causal effect estimation strategies derived from the counterfactual logic and that have been developed the last ten years.

4.1 Mature and Nascent Theories

When the constructs, their links and the underlying logic are well specified – the mature theory situation according to Edmondson and McManus (2007; cf. Zahra, 2007) – it follows directly from theory what variables need to be included and how their relationships should be modeled. If the theory is well established its boundary conditions should also be well established, meaning that the theory will also indicate what contexts or samples should be studied. In the extreme case all heterogeneity problems and their solutions can be derived directly from the theory. At the other extreme Edmondson and McManus (2007) put the 'nascent' theory situation. When the theoretical knowledge of the phenomenon is rudimentary or non-existent, increased familiarization with the phenomenon through qualitative work is likely to be needed in order to even begin building an understanding of what problems of unobserved and causal heterogeneity might exist.

We would argue that in most entrepreneurship research situations the starting point is somewhere between these two extremes (cf. Zahra, 2007). The challenge is then to systematically tease out what heterogeneity exists and determine which aspects of it are methodologically troublesome and theoretically interesting, respectively. It has been argued that "Ideal theory tests should only include the variables in the theory" (Shugan, 2006: 203). Hence, variance deriving from sources outside the theoretical domain needs to be excluded or controlled for because otherwise it leads to biased estimates of causal relationships and hampers theory development. We therefore need tools that allow us to generate theory about causal effects and tools that allows to transform theory is to effective research designs where twe can make the best use of our current knowledge knowing that it is only partial.

4.2 Identifying Sources and Specific Effects of Heterogeneity

Counterfactual argumentation or the potential outcome model is an increasingly popular way to generate theory and research designs that are better able to establish causality (Pearl, 2000). Counterfactual argumentation is a logic of inference that plays a central role in establishing causality in various fields such as political science (Fearon, 1991), history (Lebow, 2000), medicine (Höfler, 2005) and economics (Heckman, 2000).

++++++++++++++++++++++++

Insert Figure 2 about here

+++++++++++++++++++++++++

As a starting point we can assume a theory making a proposition about an X→Y relationship where we would like to estimate the size of the causal effect. A counterfactual argument is based around the question if a specific event E would had happened if the cause C would have not been present. That is, "if it had been the case that c (or not C), it would have been the case that E (or not E)": Counterfactual make claims about events that did not occur. Such propositions play an essential and elementary role in the efforts of the social sciences to assess their hypotheses about the causes of the phenomena they study. The counterfactual argumentation becomes especially important when we no longer can rely on experimental design (Pearl, 2000).

Defining causality when the causes are interrelated, partially unknown, perhaps immeasurable, and data is mostly based on observation is less straightforward and becomes a major achievement. The use of counterfactuals in theory and hypothesis testing is a way to mitigate this problem. The use of counterfactual arguments is important because they allow us to better handle the problem that many models can explain the same data, and that is the formal arguments in the model that decide causality. Differently stated, in social sciences we can observe the same causal patterns, but we can imagine different interpretations to this causal patterns. The logic of inference that is counterfactuals allows us to more clearly state different theoretical explanations and under what conditions they are likely (and not) to explain what we observe. It is through its use of logic that we can say what variables need to be present to fully examine the causality argument or treatment effect and what variables can be excluded to create a parsimonious models. The use of counterfactuals makes a clear link between the nature of theory, model and empirical testing and they are mutually dependent.

The main advantage with the counterfactual arguments is that the researcher in explaining why some particular event E occurred cannot help to explain why E occurred rather than another possible outcome or outcomes. The researcher need to be clear on what the specific cause is, why it should affect the particular outcome and what would happen in the absence of the cause being present. The later is the counterfactual argument. The counterfactual argument is only as compelling as the logic and "evidence" offered by the researcher to verify the links between the hypothesized antecedent and its expected consequences (Lebow, 2000). The counterfactual not only develops a logic of argumentation in the theory, it also leads to imagine a research strategy that provides the "empirical" confirmation for a causal hypothesis.

Let us assume that we want to test the following hypothesis: "C is a cause of event E". There are two complementary approaches to test this empirically. The first approach is to imagine that C has been absent and ask whether E could have (or might have) occurred in that counterfactual case. This leads to four different potential outcomes depicted in Figure 2 that needs to be examined in our models. For example, we can imagine a theory suggesting that unemployment increase the likelihood of starting a businessbecause those faced with unemployment have reason to seek alternative ways to provide for themselves. Quadrants I and III denote the cases that accord with this explanation– when X (employment status) changes, Y (probability of creating a start-up) changes as well (Quadrant I). When there is no change in X, no change in Y is observed (Quadrant III) . Quadrants II and IV constitute the 'counterfactual' cases. In quadrant II the question is under what conditions the proposed relationship might not hold, i.e., why does getting unemployed not lead to an increased probability engagement in self-employment. Quadrant IV depicts the final outcome: why do business start-ups while not being unemployed? Hence, we constructs a space of possible outcome with related probabilities. Theory is there to explain under which circumstances and conditions some people are more likely to end up in one quadrant than another. Basically what are the sources of heteroegeniety we need to deal with to test our hypothesis that unemployment leads to the creation of new start-ups.

Whetten's (1989) analysis of the components of theory is one possible guide to identifying sources of heterogeneity problems. Referring back to our definition of theory, in Whetten's exposition constructs constitute the *What?* of theory, whereas propositions concern the *How?* The coherent logic he associates with *Why?* whereas the *Who?, When?* and *Where?* concern the assumptions and boundaries of the theory's applicability. Although Whetten (1989) does not think of the *Who?*, *When?* and *Where?* as the likeliest candidates for strong theoretical contributions we would argue these questions provide a good starting point for to determine what heterogeneity might exist and determining what aspects of it to try to include and exclude, respectively, in the design. We hold this view because boundary conditions of theory are often not satisfactorily specified (Davis et al., 2007).

The questions *Who?, When?* or *Where?* are largely questions about the role of context, i.e., under what contextual conditions theoretical relationships are likely to hold or vary. They therefore not only inform about sources of heterogeneity, but also on the possible effects of that same heterogeneity. Johns (2006) provides a very useful discussion of what context does to empirical relationships. In short, context a) restricts range; b) affects base rates; c) changes causal direction; d) reverses signs; e) prompts curvilinear effects, and f) tips precarious relationships. This is largely about causal heterogeneity across contexts. Johns (2006) also adds a category g) 'threatens validity', which may seem superfluous as all of the above also threaten various aspects of validity. However, when the meaning or conceptual relevance of a construct itself – a *What?* of the theory – varies, we have a case of lacking 'construct equivalence' (Singh, 1995); i.e., uneven validity on the theoretical level. Contextual consequences in terms of construct inequivalence as well as causal heterogeneity highlight the need to carefully consider context in theorizing, and therefore also in design and analysis strategy.

Applying Whetten's (1989) and Johns' (2006) notions to our hypothetical case about unemployment and business start-ups several quadrant II possibilities present themselves. In a society or social stratum where most individuals are affluent by birth the construct 'unemployed' would not be equivalent with the same notion in mainstream societies. Where the institutional framework includes generous unemployment benefits and/or high bureaucratic barriers to firm formation, the relationship could be weak or non-existent. For people close to retirement age the response to lay-offs may more rarely be to set up their own business. Importantly, if the theoretically focused variable is correlated with another variable that has a negative effect on Y, the positive effect of X will not necessarily appear in empirical estimation. This would be the case here if the economic conditions that increase unemployment and therefore increase necessity-based entrepreneurship at the same time reduce opportunity-based entrepreneurship via decreased market demand (cf. Wennekers, Stel, Thurik, & Reynolds, 2005; Hamilton, 1989). These added insights into the possible nature of the relationship have consequences for sampling, variable inclusion, and modeling as further discussed in later sections. It should also be clear from the example that the counterfactual mental gymnastics has potential for better defining the boundary conditions of the theory or for enriching it with new contingencies., i.e., the exercise has repercussions on more fundamental aspects of the theory because, in Whetten's words "theorists need to understand why this anomaly exists, so that they can revise the *How* and the *What* of the model to accommodate this new information” (Whetten, 1989: 493). We would argue that the repercussions extend to the *Why?* question as well .[2]

Obviously, like in most social research we are dealing with a phenomenon that has many possible causes. This is the root of the problem of unobserved heterogeneity. These other causes pose problems when they are correlated with the cause we take a theoretical interest in. Therefore, the essence of systematic search for other causes is to find those correlated variables that either need to be included in the empirical design or made uncorrelated with X (or, more correctly, with *X’*) via constant-holding, randomization, or matching. If they are genuinely non-measurable, we may still want to assess the potential bias they cause through indirect means. Again, this will be further discussed in sections to follow.

4.3 Strategies for Causal Effect Estimation

It should now be clear that we are more interested in identification than in statistical inference. We have made two statements so far. First, entrepreneurship theories are in general not very mature. This means that we have limited knowledge of the different causal mechanisms that we are interested of. Second, we have stated that counterfactual thinking is a useful tool to think about how causality can be inferred from observational data. Our third argument is derived from the second as we here introduce the seminal work of Judea Pearl (2000), who developed a set of rules for representing causal relationships with graph theory. We do not aim to present a general introduction to his work. We only aim at introducing the most basic elements of his work in order to point towards the enormous strength of his work to help us develop better strategies for causal effect strategies.

Pearl (2000) use graph theory instead a mathematical notations because it provides a better understanding of the causal relationship to be studied, as well as it provides some

[2] For example, it is conceivable that an idealistic environmentalist currently in gainful employment and a profit-maximizing entrepreneur currently active in another industry both respond to a given institutional change by shifting from what they did previously to trying to launch similar, environment-saving new ventures.

important tools to decide what variables to include in the model and why. The idea is that we have a universe of variables that affect X and Y that be represented graphically. Given this graphic representation of the causal structure which variables must we observe and then use in the data analysis to estimate the size of the causal effect of X on Y?

The use of graph theory leads to two important advantages that allows us to better answer to the above question. First, his framework is completely nonparametric. It is therefore not necessary to specify the nature of the causal relationship between X and Y (linear, quadratic or something else). X just causes Y. This eliminates the Achille's heel of traditional path analysis which needs to make this kind of assumption. Causal models therefore become easier to handle. Second, Pearl shows that exist three basic strategies to identify a causal effect. They are: conditioning on variables that block all back doors (e.g., stratification, matching, weighting, regression), conditioning on variables that allow for estimation by a mechanism (mediation analysis), and estimating a causal effect by an instrumental variable that is an exogenous cause to the cause.

The first strategy is to eliminate sources of unwanted heterogeneity by invoking Pearl's back door criterion using available theory and measure to control for these sources (causal heterogeneity and differential validity). That is, all back-door paths from the causal variable to the outcome variable are controlled for. This strategy is called simple conditioning. The second strategy of mediation is to establish an isolated and exhaustive mechanism that relates the causal variable (X) to the outcome variable (Y) and then calculate the causal effect as it propagates through the mechanism. The third strategy is to use an exogenous instrumental variable to isolate covariation in the causal and outcome variables (Morgan & Winship (2007).An important decision criteria in how to model and test the causal relationship are the sources of heterogeneity . The most well-known and used strategies is simple conditioning. However, the point is that the use of graph theory combined with current knowledge allows the researchers to better estimate what variables to exclude and include and why, and consequently which are the best design, measurement and analytical approaches to follow.

In this section we have discussed how theory and theorizing can be used to deal with heterogeneity problems; in particular the problems of unobserved heterogeneity and causal heterogeneity. In Table 1 we summarize our advice.

+++++++++++++++++++++++++++

Insert Table 1 about here!

+++++++++++++++++++++++++++

5 DEALING WITH HETEROGENEITY IN DESIGN

It should be clear from the many sources and forms of heterogeneity discussed so far that the 'obvious' solution to this heterogeneity problem – to include all influential variables, measure them correctly and find the right model for all of their interrelationships – simply is not achievable. Neither may it be desirable, in the interest of parsimony and simplicity. A reason to prefer simpler theories is that such theories are more constraining and thus more falsifiable. A parsimonious model is not over fitted, which increases its credibility if it accords with the data (Pearl, 2000; Popper, 1959). Hence, design aims at creating a situation where the influence of one particular variable or a limited set of variables can be studied without the potentially disturbing influence of other variables. We discuss three strategies for approaching that ideal: 1) applying experimentation and other controlled approaches using 'artificial' data; 2) exploiting experiment-like real world situations, and 3) reducing heterogeneity through sampling from one or more sub-populations characterized by relatively high internal homogeneity. These three research designs are example on how to conditioning on the variables that blocks the back doors of the casual relation to be estimated. We here assume a good theoretical knowledge and the availability of good measures. When unobserved heterogeneity is a problem, then the use of longitudinal designs allow for an important safeguard. Longitudinal designs are discussed in some details in the section on related analysis techniques.

5.1 Experimentation and Other Laboratory Research Methods

Reducing heterogeneity lies at the heart of experimental research. Through manipulation of the focal explanatory variable(s) (‘treatments’) and randomization or constant-holding of other influences the experimenter can eliminate the problem of unobserved heterogeneity. In practice, experimentation is likely to reduce problems of causal heterogeneity and ununeven validity as well because of the reduced complexity of the research setting. Yet, the use of experiments has been limited in entrepreneurship research. Chandler and Lyon, (2001) and Bouckenooghe et al. (2007) both report 4 percent of the studies reviewed used an experimental approach with perhaps as little as 1 percent being based on ‘laboratory’ experiments.

With some creativity a range of entrepreneurship issues can be addressed through experiments and other methods using laboratory control (see Baron, 2006; Baron & Ward, 2004; Schade, 2005). For example, Dean Shepherd has championed empirical work using conjoint analysis and other experimental approaches to address otherwise hard-to-study issues such as opportunity recognition and effects of emotions (Brundin, Patzelt, & Shepherd, 2008; Gregoire, Barr, & Shepherd, forthcoming; Shepherd & DeTienne, 2005; Shepherd & Zacharakis, 1997). Similarly, Gustafsson (2006) compared expert and novice entrepreneurs through experimentation, largely supporting Hammond’s (1987) theoretical proposition that experts have the ability to adapt their decision-making style to the level of uncertainty associated with a particular venture, while novices tend to apply the same approach regardless of task characteristics. These experimental studies have in common that they produce relatively clear answers that are likely to be replicable. Further, Sarasvathy (2001, 2008) developed Effectuation Theory through laboratory work – non-experimental but retaining many of the heterogeneity reducing characteristics of experimental work as she had expert entrepreneurs think aloud about the same opportunity rather than studying each pursuing their own, idiosyncratic venture.

Simulations offer another type of controlled context for developing and testing theoretical ideas. According to Davis et al., (2007: 483) "simulation is especially useful for theory development when the focal phenomena involve multiple and interacting processes, time delays, or other nonlinear effects such as feedback loops and thresholds" – a characterization we would argue fits well with the reality of entrepreneurship. The studies by March (1991) and Nelson and Winter (1982) are famous examples of how this approach has been applied to topics that are entrepreneurship-related. A 'narrower' example close to the core of entrepreneurship is provided by Fiet, Piskounov and Patel (2005). We would encourage increased use of experimentation and simulation in entrepreneurship research, at least as steps towards ascertaining internal validity in a 'full cycle' approach to building evidence. In such an approach, the researcher aims at first identifying relevant problems in the field and then reducing the characteristics of the problem so that theoretical relationships can be tested in a 'laboratory' setting. Once affirmative evidence has been achieved under such controlled conditions verification in more real-life like setting can be sought (Cialdini, 1980; Chatman & Flynn, 2005). This approach can help avoid lack of progress in research on new questions arising from conflicting results which in turn are due to overwhelming heterogeneity problems.

5.2 Exploiting Experiment-Like Situations

For practical or ethical reasons most entrepreneurship research problems cannot be studied experimentally. 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 were it a uniform, permanent institutional change rather than an experiment. That is, for the most part entrepreneurship researchers have to rely on observational studies. These differ from experiments in that the investigator cannot control the assignment of the treatments to subjects (Rosenbaum, 2002). One partial remedy – in particular to unobserved heterogeneity problems – is then to work on a well defined population that exogenously receives the same treatment. Under such circumstances members of the population cannot choose whether or not they receive the treatment, but they can choose how to respond to it. Hence, there is no problem of self-selection (Heckman, 1979; 2000).

One elegant example that emphasizes how a group of dissimilar entities react to the same 'treatment' is Shane's (2000) study of 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' (all of whom came from the same university) the risk of unmeasured heterogeneity distorting the results could be reduced. The study provides compelling evidence that prior knowledge is very important in determining what specific business opportunities entrepreneurs will discover and/or succeed with under otherwise equal conditions.

Tragic as they are, natural disasters like Hurricane Katrina and the Boxing Day Tsunami also provide entrepreneurship research opportunities (see Dickson & Kangaraarachchi, 2006; Runyan, 2006). For example, comparisons of entrepreneurs (business owners or founders) with others (managers or general population) typically confound factors that make people *engage* in entrepreneurial endeavors with those factors that make them *persist* and *succeed,* respectively, at such tasks (Davidsson, 2004: 70). Post-disaster situations present a cleaner context for addressing the specific issue of entrepreneurial persistence. Galbraith and Stiles (2006) present some results on that particular issue.

Natural situations where cases vary on one variable that can be assumed to be uncorrelated to other variables, similarly to randomization in experiments, are a variation on this theme. An example here is research on the impact of windfall gains on entry into self-employment. This research is interested in how liquidity constraints hamper entry into self-employment. Lottery wins are randomly distributed among participants in the lottery, so there is no problem of self selection and the researchers can argue that the treatment effect is exogenous to the model. Lindh and Ohlson (1996) and Taylor (2001) both found that winning at the lottery increases the probability of entering into self-employment. The significance of this finding should be understood in the context of much other – and less controlled – observational research on entrepreneurial propensity finding a surprising absence of effects of financial variables (Davidsson, 2006; Kim, Aldrich, & Keister, 2006).

5.3 Narrow Sampling

While natural experiments can be creatively and opportunistically capitalized on for shedding light on some issues, they are unlikely to serve the purpose of informing the particular issues the researcher already had an interest in. When experiments and experiment-like designs are not possible the researcher can reduce bias due to unobserved heterogeneity by using a narrow sample like a single industry, a narrow age cohort or size band, or a combination of these. The main reason for this is that a range of variables Z that would influence Y in a more broadly based sample will be constants or near constants in the more restricted context, thus not contributing to variance in Y. In addition, a less complex empirical setting facilitates familiarization with the studied phenomenon, which arguably reduces the risk that relationships are misspecified for parts of the studies population, i.e., the problem of