The Survivor Principle Meets Corporate Diversification

Lasse B. Lien

TABLE OF CONTENTS

FOREWORD

Writing a thesis generates two things: learning and debts. Others shall be the judges of the amount of learning produced, but I know for a fact that the debts accumulated are substantial. So while the rest of the thesis is about learning, I devote this section to debts. In trying to list my debts I shall make no attempt to rank them by size or any other criterion, since it is hard to find a scale that would apply to both the time and attention I have stolen from Hege and my two sons, as well as to the attention and advice I have received from colleagues. What I shall do in the following is therefore merely to document my debts, and hopefully the future will bring chances to repay them.

I start by noting that I owe much to the chairman of my dissertation committee, Christine B. Meyer. I thank her for continued support, and for having faith when progress was slow (for example during the 1.5 years it took to get the data I wanted/needed). I also thank her for many valuable discussions before this thesis was written, during its writing, and hopefully many more to come now that it has been completed. Sadly, Christine was absent during some of the critical stages of writing. For reasons incomprehensible to me Christine chose to serve as a junior minister in the Norwegian cabinet over following the day to day progress of my work. I can only hope that the Norwegian electorate will appreciate this monumental sacrifice. This leads me to the person who has influenced the final result the most, Professor Nicolai J. Foss at Copenhagen Business School. He has influenced my work on many levels. Most generally through my admiration for his own writing, but more directly for two other reasons. One being that he accepted to let me come to CBS for 6 months, a period during which the basic ideas in this thesis were developed (after he showed me the folly of some other ideas I was pursuing at the time). The other is of course that in Christine's absence he has been the de facto principal advisor. This does not mean that he can be held accountable for the many

shortcomings I am sure still exist, but it does mean that without his insightful comments and advice the shortcomings would have been bigger and more numerous.

Nicolai also introduced me to Peter G. Klein of the University of Missouri, without whom this thesis could not have been produced. Peter managed to get me the data I needed. In addition Peter has become a close collaborator, and a person who has engaged me in some extremely valuable discussions, in particular with respect to the first paper. I'm happy knowing that our collaboration will continue also with respect to the three other papers, and all four of them will eventually be published jointly. Peter is not only one of the smartest persons I have ever met, he is also extremely nice to interact with, both via nightly E-mails and tête-à-tête.

A long time collaborator who has also influenced this work is Erik W. Jakobsen. His influence is both direct through the valuable comments he has given along the way, but also indirect, because discussions with him on a wide variety of intellectual subjects have profoundly shaped (and sharpened) my way of thinking and writing.

The last member of my dissertation committee is Sven Haugland, who also supplied valuable comments to each of the papers. In fact, just as I am writing these very words, more comments from Sven are arriving on E-mail. I thank him for all of them.

Another extremely important contributor is Olav Kvitastein. Completing the analyses contained in this thesis required overcoming some rather difficult programming challenges in SPSS. Olav Kvitastein taught me about the mysteries of creating syntaxes in SPSS, and when I at one point encountered a problem I didn't know how to solve, Olav found a way to solve it. This particular problem was so complex that even with his enormous skills in this area it took him about six weeks to solve it, so we are not talking about small favors here.[1] Fortunately for me, I think cracking this problem became a question of professional pride for him, something I cynically exploited.

---

[1] There is probably no one more knowledgeable about SPSS programming in Norway than Olav Kvitastein.

KEY CONCEPTS

The Survivor Principle

> The survivor principle is the assumption that the quest for economic profit combined with selection forces in competitive markets. ensures that a sample of firms in competitive markets will reflect choices and behaviors that are efficient. We refer to paper number one (p. 27) for a more thorough discussion of the origins. uses and versions of the survivor principle.

Corporate Diversification

> Corporate diversification in the sense used here is the question of which industries a diversified firm should combine within its portfolio. and the study of the determinants and consequences of such decisions. We refer to paper number two (p. 61) for a more comprehensive introduction to key issues in this field of research.

INTRODUCTION

1. Research Questions

The present thesis is about the relationship between the survivor principle and corporate diversification. This relationship is explored in two different ways. One is about what corporate diversification can do for the survivor principle; the other is about what the survivor principle can do for corporate diversification.

Let us first focus on what corporate diversification can do for the survivor principle (henceforth: SP). The SP, although widely used as an assumption in empirical tests, does not itself rest on a firm empirical footing. In fact, there does not seem to exist any explicit attempts to empirically test the SP. This is both problematic - and of considerable importance - because it means that a large part of existing empirical knowledge relies crucially on an untested assumption - namely that the SP is valid. Accordingly, what we suggest corporate diversification (and this thesis) has to offer the SP is a first direct test of its empirical validity. Although one empirical test alone cannot eliminate this concern, not in terms of relying on the SP in research on corporate diversification, and certainly not in other fields of research, it does represent a first attempt at falsification. Hence, it constitutes a first step in the direction of establishing an empirical basis for employing the SP in empirical studies. The first research question is therefore the following:

Does data from corporate diversification support the validity of empirical strategies that rely on the survivor principle?

Now, let us turn to what the SP can offer research on corporate diversification. Assuming that the SP is not demonstrably invalid, it may be potentially useful for overcoming one of the key challenges in research on corporate diversification: how to empirically capture the degree of relatedness between businesses in a corporate portfolio. In research on corporate diversification (within the field of strategic management) no other independent variable has been given the theoretical and empirical attention awarded to the issue of relatedness (Chatterjee and Wernerfelt 1991; Robins and Wiersma, 1995, 2003). Yet there is widespread disenchantment with the lack of convergence between theoretical predictions and empirical findings (Hoskisson and Hitt, 1990; Markides and Williamson, 1994, 1996; Reed and Luffman, 1986). Under normal circumstances such a lack of consistent empirical support for theoretical predictions would presumably build a pressure towards revisions of theory. In research on corporate diversification the basic theoretical predictions have to a surprising degree withstood the onslaught from poor empirical performance, because there seems to be wide agreement that the measurement procedures used to capture relatedness suffer from serious deficiencies (Hoskisson and Hitt 1990; Markides and Williamson 1994; Robins and Wiersma, 1995). So instead of focusing efforts on revisions of theory,[2] researchers have ventured on a search for better ways to measure relatedness (Fan and Lang, 2000; Farjoun, 1994; Markides and Williamson, 1996; Robins and Wiersma, 1995; Silverman, 1999). This behavior indicates that the community of research seems to put relatively more faith in the soundness of the basic theoretical arguments than in the empirical findings in this area of research.

This is where the SP may contribute to the research on corporate diversification. If valid, the SP may be used to derive empirical measures of relatedness. However, if the SP is to be useful for this purpose - it must be able to demonstrably outperform the conventional way of

---

[2] Certainly this is not to say that there is no theoretical work being done, but rather to point out that the mixed empirical findings have not led to the profound theoretical revisions one might have expected. Instead theoretical developments seem mainly to build on and extend existing theory.

measuring relatedness.[3] Our contribution is to evaluate whether this seems to be the case. Given the centrality of the relatedness variable in research on corporate diversification, and the noted dissatisfaction with existing measures, such a contribution should be of some interest.[4] Particularly because the survivor-based approach is profoundly different from all the alternatives. The second research question is therefore the following:

Do relatedness measures based on the SP outperform the conventional measurement procedures?

The reader may be entertaining an uneasy feeling of tautological reasoning at this stage. The arguments above beg the question of whether the answer to the second question is not implied by the first. The answer to this is no. Support (or lack of falsification) for the SP as an empirical strategy does not necessarily imply that relatedness measures based on the SP represent a demonstrable improvement over conventional measures. The reverse, on the other hand is impossible. If the SP is falsified as an empirical strategy for theory testing, it is obviously not valid for deriving empirical measures of relatedness either.[5] However, what is important to note at this stage is that the answer to the second question does not follow by implication from the first. We now turn to describing each paper and the relationship between them in greater detail.

---

[3] The conventional way of measuring relatedness is to use distances in the SIC-system as a proxy for relatedness. However the literature contains a plethora of other measures, including technology flows, input ratios, human resource profiles, commodity flows, etc. We refer to paper number two for an overview of the various measurement procedures.

[4] An affirmative answer to this question may also suggest that the SP can be used to overcome other measurement problems in other areas of inquiry.

[5] However one may experience a loss of excitement with respect to the answer to the first question, since if the answer was no, there would be no sense in asking the second.

2.   The Papers

The four papers that represent the body of this dissertation can be seen as constituting a 1+3 structure. This 1+3 structure refers to the two research questions discussed in the preceding paragraph, where one paper addresses the first question, and subsequently three papers than in combination seeks build an answer to the second question. Having said this, each of the three papers that addresses the second question are indeed independent works, but they are cumulative in the sense that each adds evidence to the overall conclusion.

2.1   Can the Survivor Principle Survive Diversification?

This paper addresses the question of whether data from corporate diversification support the validity of empirical strategies that rely on the survivor principle. In other words, the SP itself is put to the test. As noted previously, there is no existing study known to the author that explicitly attempts to subject the SP to a falsification test. This is problematic because a large number of empirical studies in the fields of economics, strategy and organizational economics rely on the validity of the SP in the empirical strategies chosen. The attractiveness of the SP in the context of theory testing is the idea that a sample of competitive firms can be assumed to display what is efficient. This involves the significant advantage that efficiency (which is always the ultimate dependent variable in studies relying on the SP) does not have to be measured. It can be assumed to be a property that dominates within a sample of competitive firms. Hence, the measurement task can be reduced to matching predictions of what constitutes efficient behavior to actual behavior.

So how can we test whether the empirical strategy of not measuring the dependent variable is valid? We start by assuming that which industries a diversified firm combines within its

corporate portfolio has efficiency consequences for the firm. Secondly, we tentatively assume that the SP works, which in the context of corporate diversification implies that those industries that are most frequently combined inside firms should on average represent more efficient combinations than those that are rarely combined. If this second assumption does not hold, the SP is not valid for corporate diversification, and if the SP were not valid here it would raise serious questions about whether it holds for any of the other areas where it is commonly used. This second assumption is what we design our test to examine.

To conduct such a test two steps are required. The first is to obtain a measure of the degree to which the behavior of competitive firms seems to indicate that a pair of industries are related (i.e. efficient to combine). Towards this end we estimate how much the frequencies of *actual* combinations of 4-digit SIC-industries deviate from what one would expect if diversification patterns were *random* (Teece et al., 1994). We take this difference to constitute a survivor-based measure of the relatedness between a given pair of industries. A large, positive difference indicates that they are closely related; zero or a negative difference indicates that they are unrelated. From this we can calculate measures of how related a given business i is to the other businesses in the portfolio of the parent. We calculate two such measures. One captures how related a given business i is to all the other businesses of the parent, another how related a given business i is to the two closest related businesses of the parent.

The second step is to examine whether the combinations that have been designated as efficient by this survivor-based procedure actually perform better than those designated as being less efficient. For this purpose we examine the probability that a given parent will exit one of its businesses. The prediction is that the probability of a business i being exited is lower the higher the score on both measures of survivor-based relatedness. A confirmation of this prediction would suggest that the behavior of firms in competitive markets does contain information about what is efficient, or more specifically: the efficiency of combining different

industries inside firms. Hence, the conclusion would be that the SP has survived this attempt at falsification. Given that a disconfirmation would make the question raised in the three following papers meaningless, we reveal no big secret when we say that our data strongly support this prediction.

2.3 A Common Introduction to the Three Remaining Papers

The three following papers all address the second research question, namely whether the SP can be used to obtain a better way of measuring relatedness. We will first make some introductory remarks that are relevant for all three papers, before we in subsequent sections give a brief introduction to each.

The question worth raising, given that the SP apparently works in the context of corporate diversification (in the sense that the combinations chosen by firms *do* contain information about the efficiency of various combinations), is whether the survivor-based approach is a superior procedure to empirically capture relatedness. The obvious way to answer this question is to examine whether it seems to outperform the conventional approach, which is to use relatedness measures based on distances in the SIC-system (Caves et al., 1980; Jaquemin and Berry, 1979).

Note that a survivor-based approach means that we let the actions of firms in competitive markets inform us (researchers) about which industries are related to which, instead of imposing some a priori view of what determines relatedness on the data. As such it represents an abdication in terms of letting the SIC-system or the researcher be the better judge of what is related to what. Instead a combination of the wisdom of local decision makers and the screening function of the competitive process assumes responsibility for the quality of the

measure. The procedure also means that we will not know what causes beneficial relatedness effects in each particular instance, and as such we will not be much wiser in terms of knowing what relatedness *is*.[6] But if proven superior to the conventional approach, we may improve our ability to examine what relatedness *does* - with respect to its effects on various dependent variables of interest (i.e. performance, growth, entry mode, financing, organizational choices, etc.).

How does one test whether the survivor-based approach to relatedness is actually superior to the SIC-based approach? One way of doing so is to identify some variable that relatedness can be expected to affect, and measure the ability of the different measures to explain variance in this variable. A poor measure should be able to explain less of the variance in such a variable than a good measure, as is shown in figures 1 and 2 below. The large square represents the total variance in the focal variable. A portion of this variation is explained by the true effect of relatedness on this variable, which is represented by a square inside the larger square. However, an imperfect measure of relatedness will only capture a portion of this true effect. The higher the quality of the relatedness measure, the larger the captured portion of the true effect, and the more of the total variance in the focal variable will be explained. This is reflected by the increase of the shaded area as we move from a poor measure (fig. 1) to a better measure (fig. 2).

---

[6] Although indirectly, the procedure may also be helpful for this purpose. One can for example let measures of survivor-based relatedness be the dependent variable, and examine hypotheses about causes of relatedness (i.e. variables that explain relatedness).
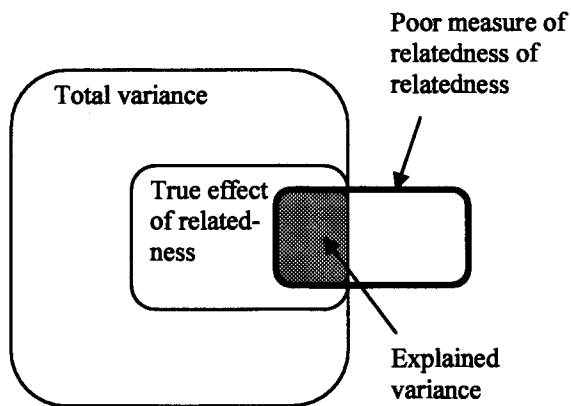
FIGURE 1: Poor measure of relatedness   FIGURE 2: Better measure of relatedness
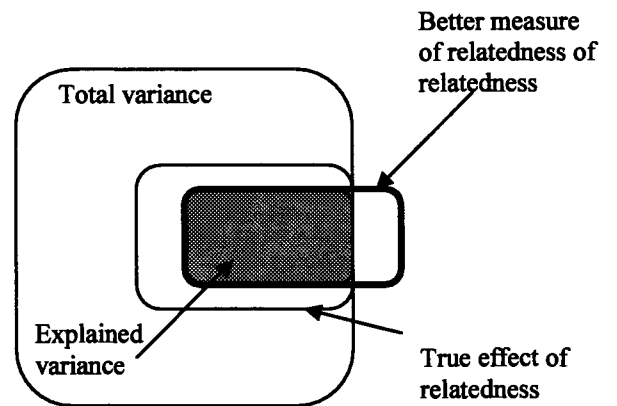
However, there is reason to be cautious in using such a logic. The reason is that a superior ability to explain variance can be spurious, that is, an effect of one or more exogenous variables that the two measures correlate differently with, and that impact the variable whose explained variance is being used for evaluation. This is depicted in Figure 3 below. Note that if the different measures correlate equally with such exogenous variables, they will not affect the relative performance of the measures, but the absolute level of explanation will of course be affected. This means that it is important to control for such exogenous variables, in particular those that can be expected to correlate differently with the different measures one is comparing.
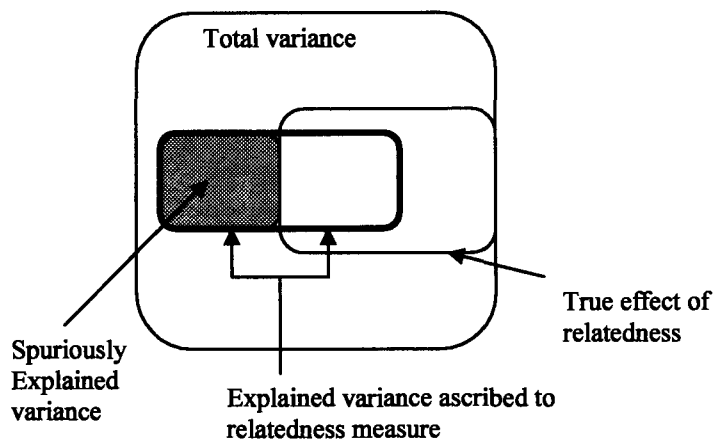
FIGURE 3: Spuriously explained variance

With these remarks in mind, let us go back to the three papers. Two variables that relatedness *can* be theoretically expected to affect are entry and exit decisions by diversified firms. Accordingly, the ability to explain the probability of entry and exit represents an opportunity to compare the performance of the survivor-based and SIC-based measures. This is what is done in paper number 2 (exit) and paper number 3 (entry). We find that the survivor-based measures explain more of the probability of both exit and entry than does the SIC-based measures.

But as noted this finding could potentially be spurious, resulting from different correlations with exogenous variables that affect entry and exit decisions. To reduce this likelihood a number of control variables are included in both papers. However, even though these control variables include the variables most frequently noted as predictors of entry and exit, the spuriousness issue is not satisfactorily dealt with by the inclusion of these variables. In particular there are two important candidates to create spuriousness that are not covered. These two are of particular importance, because they can indeed be expected to influence the relatedness measures we are comparing differently.

We are referring to the tendency towards mimetic- or herd behavior, and the pursuit and exploitation of mutual forbearance through multipoint competition. The survivor-based measures, because of the way they are constructed, are more likely to capture such influences than are SIC-based measures. Accordingly, the noted superior ability to explain exit and entry may be an artifact thereof (i.e. not a superior ability to measure relatedness). We therefore devote the fourth, and final paper to examine whether the superior performance of the survivor-based measures found in paper number 2 and 3 can be explained by these phenomena. Our findings on this issue strongly indicate that they cannot.

We now move on to some brief remarks on each of the three papers.

2.4   Yet Another Way of Measuring Relatedness - This One: Let Competition Do It!

This paper compares survivor- and SIC-based measures of relatedness in terms of explaining the probability of exit. As discussed in the previous section, the rationale is that a better measure will explain the probability of exit better - but notably we need to be cautious about the possibility that a superior ability to explain exit is spurious. We shall now briefly describe how the paper makes a comparison of the two measures, and which safeguards are built into the paper to reduce the risk of spuriousness affecting the conclusions.

We started by registering which industries a diversified firm participated in, in 1981. Our focus is on firms that had exited some industries and remained in others by the year 1985. Firms that were liquidated or sold in their entirety between 1981 and 1985 were excluded, since such actions do not reveal information about the merits of combining different industries inside firms, while exiting some businesses and keeping others do. From among such firms a sample of 1191 businesses were chosen, of which 593 were exits and 598 were non-exits. A logistic regression analysis was then conducted to evaluate which of the

relatedness measures could best explain the probability that a given parent exited a given industry. The analysis found that all measures based on the survivor approach significantly outperformed all measures based on the SIC-approach.

Spuriousness may potentially affect such an analysis in two ways. Both relatedness measures may correlate (equally) with some exogenous variable. This will have the effect of distorting the absolute level of explanation, but not the relative level of explanation. Secondly, and for the present purpose even more damaging, an exogenous variable may correlate differently with the different measures. This will both affect the absolute level of explanation and the relative level of explanation. To reduce this threat to the validity of the conclusions we included several control variables that represent the most frequently noted influences on exit decisions. On the industry level these are: industry growth, industry concentration and industry profitability. On the firm level they are: market share in the focal industry, parent size (sales), parent leverage and parent liquidity. Controlling for these variables, the conclusion remains that all measures based on the survivor approach significantly outperformed all measures based on the SIC-approach.

2.5   Relatedness and Patterns of Diversification: A Survivor-Based Approach

Based on the assumption that relatedness is relevant for entry decisions,[7] this paper compares survivor- and SIC-based measures of relatedness in terms of explaining the probability that a given parent enters a given industry. In a similar manner to the previous paper, we developed a sample of 1202 entries and 1176 non-entries made by diversified firms between[8] 1981 and 1985. Next we conducted a logistic regression analysis to evaluate which of the relatedness

---

[7] I.e. we assume that ceteris paribus, a diversified firm is more likely to enter an industry that is related to its existing businesses. There is in fact a number of studies supporting such an assumption (Farjoun 1994, Montgomery and Hariharan 1991, Silverman, 1999)

[8] Between means that we identify businesses a parent was not active in, in 1981, which it may or may not have entered by 1985.

measures could best explain the probability that a given parent entered a given industry. This analysis, like the one on exit data, found that all measures based on the survivor approach significantly outperformed all measures based on the SIC-approach.

In order to reduce the threat to this conclusion from spuriousness, we controlled for a number of industry and parent firm variables that may affect entry decisions. These include industry growth, industry concentration, industry profitability, parent size (sales), and parent diversity (the number of industries participated in 1981).

In sum, the findings in this paper add to the evidence from the previous paper by validating the superiority of the survivor-based measures against yet another variable.

2.6   Survivor-based Measures of Relatedness: Two Alternative Interpretations

This paper examines two alternative interpretations of the findings in the two previous papers. The motivation is essentially that there is a deficiency in the controls for spuriousness in these papers, and that the relevant omissions are of particular importance since we are talking about influences that could inflate the explanatory power of the survivor-based measures relative to the SIC-based measures.

The first is associated with the impact of herd behavior, which refers to a tendency among decision makers to suppress their private information, and follow the herd. This could be because a bad decision is not as damaging for a manager's reputation when others make the same mistake (Scharfstein and Stein, 1990). Or it could be because managers believe that the actions of others reflect some private information that others have (i.e. if everybody does it, it must be clever), however such tendencies means that the actions of each decision maker becomes less informative to the others (Banerjee, 1992).

The crucial point here is that the survivor-based measures can be expected to capture such tendencies better than the SIC-based measures. The reason is simply that the survivor-based approach involves explicitly measuring which combinations firms in the same industries as the focal firm have chosen in a period recently preceding the decision period. These are presumably the firms the focal firm would be herding after. The SIC-based approach, on the other hand, is based on standardized distances in the SIC-system, and it plainly does not reflect what "others are doing" nearly as well. There are for example numerous examples of industries that are close in the SIC-system, but are never combined inside firms, and the reverse: that firms are distant in the SIC-system, but frequently combined.

To examine the possibility that herd behavior accounts for the superior performance of the survivor-based measures we developed a sample of 229 entries made between 1981 and 1983, and compared the ability of the two measurement approaches to predict the probability that an entry decision was reversed by 1987. The underlying logic is here that while herd behavior may influence entry decisions, once entry has occurred competitive forces and economic reality sets in and begins it work of screening the good decisions from the bad. If SIC- and survivor-based measures capture true relatedness equally well, we would expect no difference between the two approaches in terms of predicting reversal of entry decisions. However, we again found that the survivor-based measures significantly outperformed the SIC-based measures, which is inconsistent with the suggestion that the previously noted superiority of the survivor-based measures will disappear when contamination from herd behavior is taken into account.

The second alternative interpretation is based on mutual forbearance through multipoint competition (Edwards 1955). This hypothesis suggests that high levels of contact between firms across markets will induce a balance of terror where competitors refrain from attacking

each other, and thereby instigate a condition of less vigorous competition than would otherwise have occurred (Karnani and Wernerfelt, 1985). Research has provided empirical support for the claim that creation and exploitation of mutual forbearance affects the behavior and patterns of diversification of diversified firms (Greve and Baum, 2001). The potential influence of this motive on entry and exit decision is especially worrying because the survivor-based approach is likely to capture such motives better than the SIC-based measures. The reason is that the survivor-based measures of relatedness are constructed from what essentially amounts to a count of frequencies of multimarket contact. The prediction is that firms will be more likely to enter industries where these counts will be high, and that they are less likely to exit those where they are high. The very same predictions would result from a mutual forbearance argument. Conversely, SIC-based measures are likely to be less contaminated by such motives, because they are not constructed on the basis of multimarket contact.

To examine the possibility that mutual forbearance accounts for the superior performance of the survivor-based measures found in the two previous papers, we split the samples used in these papers into two equally sized subsamples: one containing the most highly concentrated industries, and one containing the least concentrated industries. The mutual forbearance motive is only plausible in industries where concentration exceeds some minimum level, and we therefore expected the survivor-based measures to outperform the SIC-based measures in the high concentration subsamples (where the mutual forbearance motive is plausible), but not in the low concentration subsamples (where the mutual forbearance motive is not plausible). Rerunning the original regressions on these subsamples, we found that the mutual forbearance interpretation was strongly contradicted. In fact the superiority of the survivor-based measures held for all subsamples, and was indeed larger in the low concentration subsamples. In other

words. the superiority of the survivor-based measures does not seem to disappear when contamination from herd behavior is taken into account.

REFERENCES

Banerjee AV. 1992. A Simple-Model of Herd Behavior. *Quarterly Journal of Economics* 107(3): 797-817.

Caves RE, Porter ME, Spence AM. 1980. *Competition in the open economy : a model applied to Canada.* Harvard University Press: Cambridge, Mass.

Chatterjee S, Wernerfelt B. 1991. The Link between Resources and Type of Diversification - Theory and Evidence. *Strategic Management Journal* 12(1): 33-48.

Edwards CD. 1955. Conglomerate bigness as a source of power, *NBER conference report Business Concentration and Economic Policy.* Princeton University Press.: Princeton.

Fan JPH, Lang LHP. 2000. The measurement of relatedness: An application to corporate diversification. *Journal of Business* 73(4): 629-660.

Farjoun M. 1994. Beyond Industry Boundaries - Human Expertise, Diversification and Resource-Related Industry Groups. *Organization Science* 5(2): 185-199.

Greve HR, Baum JAC. 2001. Introduction: A multiunit, multimarket world, *Advances in Strategic Management,* Vol. 18: 1-28.

Hoskisson R, Hitt M. 1990. Antecedents and Performance Outcomes of Diversification: Review and Critique of Theoretical Perspectives. *Journal of Management* 16: 461-509.

Jaquemin A, Berry C. 1979. Entropy Measure of Diversification and Corporate Growth. *Journal of Industrial Economics* 27: 359-379.

Karnani A, Wernerfelt B. 1985. Multiple Point Competition. *Strategic Management Journal* 6(1): 87-96.

Markides CC, Williamson PJ. 1994. Related Diversification, Core Competence and Corporate Performance. *Strategic Management Journal* 15: 149-165.

Markides CC, Williamson PJ. 1996. Corporate Diversification and Organizational Structure: A Resource-Based View. *Academy of Management Journal* 39: 340-367.

Montgomery CA, Hariharan S. 1991. Diversified Expansion by Large Established Firms. *Journal of Economic Behavior & Organization* 15(1): 71-89.

Reed R, Luffman GA. 1986. Diversification: The Growing Confusion. *Strategic Management Journal* 7: 29-36.

Robins J, Wiersema MF. 1995. A Resource-Based Approach to the Multibusiness Firm - Empirical-Analysis of Portfolio Interrelationships and Corporate Financial Performance. *Strategic Management Journal* 16(4): 277-299.

Robins JA, Wiersema MF. 2003. The measurement of corporate portfolio strategy: Analysis of the content validity of related diversification indexes. *Strategic Management Journal* 24(1): 39-59.

Scharfstein DS, Stein JC. 1990. Herd Behavior and Investment. *American Economic Review* 80(3): 465-479.

Silverman BS. 1999. Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. *Management Science* 45(8): 1109-1124.

Teece DJ, Rumelt R, Dosi G, Winter S. 1994. Understanding Corporate Coherence - Theory and Evidence. *Journal of Economic Behavior & Organization* 23(1): 1-30.

Can the Survivor Principle Survive Diversification?

Lasse B. Lien

Department of Strategy and Management

Norwegian School of Economics and Business Administration

Breiviksveien 40

N-5045 Bergen

Norway

lasse.lien@nhh.no

Tel: +47 55 959 726

Abstract: The survivor principle holds that hypotheses about efficient firm behavior can be tested by observing how firms actually behave in competitive markets. This principle underlies much of the empirical work in organizational economics and strategic management. The validity of the survivor principle itself, and subsequently the empirical work that relies on it, boils down to an empirical question: To what extent do competitive markets actually display what is efficient? Despite the centrality of this question for the accumulation of knowledge, and despite the extensive theoretical discourse surrounding it, we have not seen any attempts to settle this issue empirically. This paper contributes an empirical test of the survivor principle within the area of corporate diversification. Our findings provide support for the validity of relying on the survivor principle in empirical tests.

1. Introduction

Much of the empirical work in organizational economics and strategic management assumes that we can infer what strategies or firm characteristics are efficient by observing what firms actually do. Do high levels of asset specificity require more hierarchical modes of governance? Regress the decision to vertically integrate on a measure of asset specificity. In what industries do the incentive effects of performance-based pay outweigh the losses from inefficient risk sharing? Simply regress the use of performance-based pay on firm and industry characteristics. In other words, to see what strategies or structures work well with what attributes, or what combinations of business decisions work well together, we look at actual behavior, assuming that markets are sufficiently competitive. Hypotheses about efficient behavior can be tested by observing which behaviors dominate in populations of competitive firms.

This assumption is often referred to as the *survivor principle*. The name was coined by Stigler (1968), but the ideas are usually credited to Alchian (1950) and Friedman (1953).[9] Alchian argued that even though economic theories about rational decision makers making efficient choices are clearly unrealistic, the predictions of such a theory are not. The quest for economic profit, combined with selection forces in competitive markets, ensures that the behavior of competitive firms will roughly approximate the substantive predictions of such a theory (Alchian, 1950, p. 211). Or, as Friedman (1953, p. 22) puts it

> [U]nless the behavior of businessmen in some way or other approximated
> behavior consistent with the maximization of returns, it seems unlikely that

---

[9] Some trace the concept back to Harrod (1938). Another early contributor is Enke (1951).

they would remain in business for long. Let the apparent immediate

determinant of business behavior be anything at all—habitual reaction,

random choice, or whatnot. Whenever this determinant happens to lead to

behavior consistent with rational and informed maximization of returns, the

business will prosper and acquire resources with which to expand; whenever it

does not, the business will tend to lose resources and can be kept in existence

only by the addition of resources from outside. The process of "natural

selection" thus helps to validate the [maximization] hypothesis—or, rather,

given natural selection, acceptance of the hypothesis can be based largely on

the judgment that it summarizes appropriately the conditions for survival.

Note that the general claim is that surviving firms will behave "efficiently," however that is

defined, not necessarily that this behavior is particularly well described by neoclassical

economics. Thus, while transaction cost economists may claim that the efficiency calculus of

neoclassical economics gives insufficient consideration to bounded rationality and

opportunism (Williamson, 1975, 1985), and resource-based theorists may claim that it

downplays factor market imperfections (Wernerfelt, 1984; Barney, 1986; Diercickx and

Cool, 1989); they may still accept a general version of the survivor principle. Williamson

(1988, p. 174), for example, notes that empirical research in transaction cost economics

"relies in a general, background way on the efficacy of competition to perform a sort between

more and less efficient modes and to shift resources in favor of the former."

Judging from its adoption in empirical studies, the survivor principle seems widely accepted.

Indeed, as argued below, it can be described as the default empirical strategy in economics

and organizational economics, and it is also common in strategic management, and several

related fields. But how well does the economic natural selection process work? Are

inefficient outcomes eliminated quickly, or with long and variable lags? How important are

industry and economy-wide characteristics such as regulation, capital-market performance,

anti-takeover amendments, and bankruptcy codes? Williamson acknowledges that the process

of transaction cost economizing, for instance, is not automatic. The efficient alignment of

transactions and governance structures

> seems plausible, especially if the relevant outcomes are those that appear over
> intervals of five and ten years rather than in the very near term. This intuition
> would nevertheless benefit from a more fully developed theory of the selection
> process. Transaction cost arguments are thus open to some of the same
> objections that evolutionary economists have made of orthodoxy (Williamson,
> 1988, p. 174).

Given its widespread (and controversial) use, it is surprising that the survivor principle itself

has not been subject to empirical testing. This paper investigates the survivor principle in the

context of exit decisions by diversified firms. We begin by assuming that which industries are

combined within a firm has consequences for the efficiency of the firm. If the survivor

principle holds, we can further assume that those pairs of industries that are most frequently

combined within firms on average represent more efficient combinations than those pairs of

industries that are rarely combined. In other words, the behavior of competitive firms can

show us which combinations are efficient. If so, a diversified firm should be significantly

more likely to exit those industries that this "survivor logic" identifies as a poor match with

other businesses in the portfolio, compared to those identified as representing a good match.

The paper proceeds as follows: Section 2 reviews the content and critique of the survivor

principle. Section 3 presents how the frequency of combinations of industries in diversified

firms can be converted to a survivor-based measure of relatedness, and from this we develop

hypotheses that test the survivor principle in the context of corporate diversification. Section

4 discusses our empirical approach. Section 5 presents our results, and section 6 concludes.

2. The survivor principle

Alchian's (1950) initial formulation of the survivor principle was conducted in a period when economic theory was under increasing attack for being based on unrealistic representations of human decision making. The critics argued that since actual decision makers did not have the information nor the processing capabilities assumed in economic models (Simon, 1947), and generally did not make decisions by way of marginal analysis (Lester, 1946), the value of conventional economic theory was questionable. Alchian made the point that even though economic models may be unrealistic as process models of how decisions are made; the outcomes predicted by economic theory were likely to be both robust and quite accurate. Two key processes would ensure this. One was that firms making negative profits would, unless some corrective measure was taken, lose resources and ultimately become extinct, while firms making positive profits would acquire resources and grow. The other was that the desire to make positive profits would provide a strong incentive for the less successful firms to imitate the more successful firms. In combination, Alchian postulated that these two forces implied that surviving firms in competitive markets would appear "as if" they were behaving in the manner described by economic theory.

Friedman (1953) took Alchian's ideas one step further (or too far, as some would argue) in two respects. The first is that while unrealistic behavioral assumptions to Alchian was something that could be compensated for by the selection forces and profit incentives of the competitive process, to Friedman unrealistic assumptions were a virtue rather than a vice. In a famous methodological essay he argued the point that building theories on realistic assumptions is not only infeasible, but also undesirable (Friedman, 1953). A good theory, according to Friedman, is a theory that explains much by little, in the sense of generating good predictive accuracy by selecting a few simplifying assumptions that removes the clutter

and detail of the real world. In Friedman's view the goal of building theories on realistic assumptions would produce theories that were mere imperfect representations of reality. He pronounced conventional economic analysis to be an example of a useful theory, not because its assumptions are realistic, but because the forces described by Alchian would ensure great predictive accuracy. Secondly, Friedman went further than Alchian with respect to the accuracy of these predictions. He claimed that the competitive process would produce outcomes consistent with optimizing behavior, while Alchian made the more modest claim that the competitive process would systematically select the best among the tested alternatives (survival of the fitter, rather than survival of the fittest).[10]

The arguments of Alchian and Friedman have been put to two different uses, one theoretical and one empirical. The theoretical use is a defense for explaining economic institutions and economic behavior on the basis of efficiency consequences; even if such theories are far from realistic accounts of the causal processes that create the phenomena of interest (Dow, 1987). Such theories are referred to as functionalistic explanations, and they dominate within economics and organizational economics. The other, and for our purposes more important use, is related to theory testing. If the survivor principle holds, a sample of competitive firms can by and large be assumed to display what is efficient. This involves the significant advantage that efficiency does not have to be measured. It can be assumed to be a property that dominates a sample of competitive firms. When testing theories that have efficiency as the ultimate dependent variable, the measurement task can be reduced to matching predictions of what constitutes efficient behavior to actual behavior.

This empirical strategy is widely adopted (though only rarely explicitly stated). We recognize it from empirical tests of transaction cost analysis, where for example the hypothesis that vertical integration is more efficient than market governance when asset specificity is high, is

---

[10] Thus it is quite possible to agree with Alchian but disagree with Friedman, a point we elaborate below.

tested by measuring whether firms actually integrate when assert specificity is high (see

Shelanski and Klein. 1995, and Boerner and Macher. 2002, for surveys of a vast empirical

literature). We also recognize it from agency theory where hypotheses about the relative

efficiency of alternative contracts are tested by measuring which contracts firms actually

employ (e.g., Anderson. 1985; Eisenhardt. 1985). And we recognize it from studies of

diversification within the strategic management literature. where for example hypotheses

about what constitutes efficient patterns of diversification are tested by measuring their

consistency with actual patterns of diversification (e.g., Farjon. 1994; Montgomery and

Hariharan. 1991; Silverman. 1999; Matsusaka. 2001). These are just a few examples of

empirical papers that rely on the survivor principle. In addition it is widely used in industrial

organization and property rights theory, and occasionally within finance and marketing. A

comprehensive review extends far beyond what can be accomplished here, but it seems safe

to conclude that use of the survivor principle is a central empirical strategy in the study of

organizations and their behavior.

Given its widespread use one may be led to believe that the survivor principle is

uncontroversial. However. it is not. Right from the outset critics have questioned its use in

both theory development and theory testing (Penrose, 1952). Regarding theory development

critics have argued that the survivor principle encourages abandoning the goal of building

theories that provide true accounts of the relevant causal processes, and that it is therefore

detrimental to scientific progress. In particular there has been a long and heated debate over

Friedman's (1953) position on the (ir)relevance of building theories on realistic assumptions

(see, for example. Blaug. 1980; Boland. 1979; Caldwell. 1980; Musgrave. 1981; Mäki. 1994).

For our purposes the critique of the survivor principle as an assumption in theory testing is

more important.

Is the selection processes in competitive markets sufficiently fast and precise to justify the assumption that surviving firms are "efficient" or at least approximately so? The answer depends on whether we focus on Friedman's "optimizing" version of the survivor principle or Alchian's "comparative efficiency" version. One important argument against the optimizing version relates to the problem of sufficient variation. Since selection only operates on the tested behavior or decisions, there must be sufficient variation so that the optimal behavior is part of the set selection operates on. If not, selection cannot produce an optimal outcome (Nelson and Winter, 1982). The idea that variation is sufficient for the optimal behavior to always be a part of the set of tested solutions appears to be quite a bold assertion. On the other hand one may claim that even if the optimal solution is not part of the initial set, entrepreneurship, incremental learning and experimentation will soon make it so. However, for selection to produce optimal outcomes in this manner, the benefits from changing towards the optimum must be continually increasing. If there is a local optimum next to a global optimum, it may be impossible to reach the global optimum from the local optimum; there could be a portion between the two points where the benefit curve is downward sloping, and hence the optimum point cannot be reached through evolution by small steps (Elster, 1989). Even if there is no such negatively sloped portion of the benefit curve, the need for a slow evolution towards the optimum would mean that the system is outside the optimum for substantial periods of time, which would of course seriously damage the descriptive accuracy of the optimizing version of the survivor principle. Thus, the idea that competition is an optimizer seems hard to accept as an accurate description of empirical reality.

Alchian's comparative efficiency version of the survivor principle, which is the version relied upon in most empirical work, is not vulnerable to this critique.[11] His argument is that

---

[11] Alchian's version is the one implicitly used in most empirical work because hypotheses are typically formulated as comparative statements about efficiency, for example between hierarchical and market governance, related and unrelated diversification, fixed or variable compensation, and the like. Empirical tests

selection will operate on the tested solutions only, and bring about a situation where the comparatively best of these dominate. He explicitly points out that this does not imply optimal solutions: "Positive profits accrue to those who are better than their competitors, even if the participants are ignorant, intelligent, skillful, etc. . . . As in a race, the award goes to the relatively fastest, even if all competitors loaf" (Alchian, 1950, p. 213). However, this more modest version of the survivor principle is not immune to some of the other critical arguments against the survivor principle. For example Winter (1964, 1971) has pointed out that because of environmental change, selection has a moving target. If environmental conditions change at a higher rate than selection and adaptation processes operate, it becomes difficult to say which environmental conditions a population is adapted to. In other words, the populations we observe today may be dominated by the solutions that were efficient in the environment of yesterday. To what extent this invalidates the survivor principle seems ultimately to be an empirical issue concerning the relative speed of environmental change vs. selection and adaptation, which cannot be decided a priori.

Another important (and related) objection concerns the multitude of decisions that affect performance. While the former point dealt with the stability over time of the conditions that determine what constitutes an efficient choice, the current point has to do with the numerous conditions that at any one time will affect performance. Given that selection operates on actual performance, and not the causes of performance (Elster, 1989), a surviving firm may contain a mixture of efficient and inefficient choices. Particularly if selection forces are weak, a firm may survive even if some of the decisions made are inefficient, as long as other decisions cause sufficient efficiency to ensure survival. Thus, a survivor-based measure of what constitutes an efficient choice will be noisy and probably noisier the less important the decision studied. By importance we here refer to the efficiency consequences of not choosing

---

of such comparative statements do not rely on the optimizing version of the survivor principle, but only that relatively more efficient outcomes will tend to be observed.

the best alternative. But again, the weight of this argument seems to be a matter that can only be settled empirically.

Furthermore, it has been pointed out that there is a feedback mechanism between market selection and environmental conditions (Hodgson, 1993). For example, if one firm makes a decision to enter a new market, and is successful, competitors may follow suit. As more firms enter the new market, profitability is pushed downward, and it is possible that what was efficient for the first mover becomes inefficient for the later movers - or even for the first mover as well. This scenario could apply to new technologies, new organizational designs, or new distribution forms as well. The challenge this presents to the survivor principle is that it is precisely widespread adoption in a population that converts what was once efficient to inefficiency. If such scenarios occur often, it is damaging to the idea that a population is dominated by the comparatively efficient solutions. Whether they do occur often is yet another matter that can only be settled empirically.

In sum, there are some important objections to using the comparative efficiency version of the survivor principle as an assumption in empirical tests. However, the fact that empirical strategies that employ the survivor principle are noisy is not a sufficient reason to discard it a priori. The question that must be asked is whether it is *noisier* than alternative available strategies. If we instead attempt to measure efficiency *directly* we will also obtain less than perfect measures. The question thus becomes one of the relative noise of different empirical strategies. If there are situations where survivor-based approaches are less noisy than available alternatives, then such strategies should continue to play a role in empirical research. But to settle this question requires that the survivor principle itself be scrutinized empirically. A logical first step in doing so seems to be to examine empirically the assumption that decisions or behaviors that occur frequently in a population of competitive

firms are *on average* more efficient than those that occur rarely. We now proceed with such a test within the area of corporate diversification.

3. A survivor-based measure of relatedness

As noted in the introduction a fundamental assumption in the present study is that which industries are combined within a firm has efficiency consequences for the firm. We shall refer to these efficiency consequences as the *relatedness* of a given industry to other industries in the portfolio of a given firm.[12] However, we will not make any attempt to specify exactly what causes such efficiency consequences, but instead we assume—in concordance with the survivor principle—that the decisions made by competitive firms can reveal the relatedness between any given pair of industries. The fundamental premise of this survivor-based approach to relatedness is thus that industries that are related will be more frequently combined in one firm. More specifically, we estimate how much the frequencies of *actual* combinations of 4-digit SIC industries deviate from what one would expect if diversification patterns were *random*. We take this difference to constitute a survivor-based measure of the relatedness between a pair of industries. This method was originally proposed and developed by Teece, Rumelt, Dosi, and Winter (1994).

Let the universe of diversified firms consist of K firms, each active in two or more of I industries. Let $C_{ik} = 1$ if firm k is active in industry i. The number of industries participated in by firm k is $m_k = \Sigma_i C_{ik}$ and the number of diversified firms present in industry i is $n_i = \Sigma_k C_{ik}$. Let J_{ij} be the number of diversified firms active in both industries i and j, such that

---

[12] Relatedness has primarily been studied at the inter-industry level, but since firms in any industry differ in their resources there may be a relatedness component that is firm specific (see, for example, Silverman, 1999). In line with most of the existing work, the focus is here on inter-industry relatedness.

$J_{ij} = \Sigma_k C_{ik} C_{jk}$. Thus J_{ij} is a count of how often industries i and j are actually combined within the same firm. J_{ij} will be larger if industries i and j are related, but will also increase with n_i and n_j. To remove the effect of the size of industries i and j, the number J_{ij} is compared with the number of expected combinations if diversification patterns were random.

The random diversification hypothesis can be operationalized as a hypergeometric situation where a sample of size n_i is drawn (without replacement) from a population of K firms. Those chosen are considered to be active in industry i. A second independent sample of size n_j is then drawn from the population the population of K firms. Those chosen are considered active in industry j. The number x_{ij} of firms active in both i and j is then a hypergeometric random variable with population K, special members n_i and sample size n_j. The distribution function for this variable is then:

$$\Pr(X_{ij} = x) = f_{hg}(x,K,n_i,n_j) = \frac{\binom{n_i}{x}\binom{K-n_i}{n_j-x}}{\binom{K}{n_j}}$$

The mean and variance of X_{ij} are:

$$\mu_{ij} = E(X_{ij}) = \frac{n_i n_j}{K}.$$

$$\sigma^2 = \mu_{ij}\left(1-\frac{n_i}{K}\right)\left(\frac{K}{K-1}\right).$$

A standardized measure of the relatedness between industries i and j is then constructed based on the difference between J_{ij} and μ_{ij} in the following fashion:

$$SR_{ij} = \frac{J_{ij} - \mu_{ij}}{\sigma_{ij}}$$

The measure SR_{ij} is thus a standardized measure of how much actual combinations exceed expected combinations under the random diversification hypothesis. With this fundamental measure in hand it is possible to establish a survivor-based measure of how related a given business in a corporate portfolio is to the other businesses in the same portfolio. Again the procedure is based on Teece et al. (1994).

Assume a diversified firm that participates in m industries. Its business in industry i has sales of s_i and survivor-based relatedness SR_{ij} with industry j. The weighted average relatedness WAR_i of the business in industry i to all other business in the firm is then defined as:

$$WAR_i = \frac{\sum_{j \neq i} SR_{ij} s_j}{\sum_{j \neq i} s_j}$$

An alternative approach does not consider how related each business is to *all* other business in the corporate portfolio, but how related each business is to its *two closest* neighboring businesses. The approach here is to rank the survivor-based measure SR_{ij} between a given industry, and all other industries in the parent portfolio. The two industries with the highest measure of SR_{ij} are considered the neighboring businesses. Let $\lambda_{ij} = 1$ for a business that is defined as a neighbor to business i, and $\lambda_{ij} = 0$ for those that are not. The weighted average relatedness of neighbors to business i is then defined by:

$$WARN_i = \frac{\sum_{j \neq i} SR_{ij} s_j \lambda_{ij}}{\sum_{j \neq i} s_j \lambda_{ij}}$$

Note that the measures WAR_i and $WARN_i$ are indeed survivor-based measures of relatedness, since they are created under the explicit assumption that the diversification decisions made by competitive firms reveal information about the efficiency of different combinations. Yet, actual diversified firms will to varying degrees contain combinations that have been designated as efficient by this procedure. The survivor principle implies that businesses

scoring low on these measures are more likely to be exited than businesses scoring high. Such businesses are likely to be associated with lower performance than those that fit well because they benefit less from economies of scope with the other businesses, and because they add complexity and governance costs to the firm (Prahalad and Bettis, 1986). Furthermore, there is a comparatively higher probability that owners with better combinations will bid more for such a business than the current owners expect it to generate (Goold, Campell, and Alexander, 1994). For these reasons we expect them to be associated with high probability of exit, whether by closure or divesture. More specifically we get the following hypotheses:

> H1: Businesses with low levels of survivor-based relatedness (WAR_i) to *all other* businesses in the corporate portfolio are more likely to be exited than businesses with high levels of survivor-based relatedness.

> H2: Businesses with low levels of survivor-based relatedness ($WARN_i$) to the closest *neighboring* businesses in the corporate portfolio are more likely to be exited than businesses with high levels of survivor-based relatedness.

If these hypotheses should fail to receive support this would presumably be either because the combination of industries in firms does not have efficiency consequences, or because the decisions of firms in competitive markets are unable to reveal which combinations are efficient. Since we hold the former to be unlikely,[13] we submit that this constitutes an empirical test of the latter - that is, a test of the survivor principle.

---

[13] The findings regarding the performance of corporate portfolios are admittedly mixed (Hoskisson and Hitt, 1990; Ramanujam and Varadarajan, 1989; Robins and Wiersma, 1995). However, by most accounts this is not because corporate diversification is irrelevant, but because relatedness is difficult to measure. Like Teece, Dosi, Rumelt, and Winter (1994), we do not impose a particular view of relatedness on the data. Instead we let the data tell us what appears to be related to what.

4. Empirical approach

This study involves two distinct empirical operations. The first operation is to develop the

fundamental survivor-based measures of relatedness SR_{ij} for any pair of industries that are

combined in a diversified firm, in order to be able to calculate the key independent variables

WAR_i and $WARN_i$. The second empirical operation is to test H1 and H2. Since the data,

samples, variables and method of analysis vary between these two empirical operations, we

discuss methodological issues separately for each.

4.1. Calculating relatedness

To calculate the survivor-based measure of relatedness SR_{ij} we used the AGSM/Trinet Large

Establishment Database (Trinet). The Trinet database contains biannual records of all U.S.

establishments[14] with more than 20 employees from 1979 to 1989[15], including data on 4-digit

SIC code, corporate ownership, and sales. By aggregating the establishments for each parent

in each 4-digit sic-code, and the different 4-digit sic-codes for each parent, and different

parents for each 4-digit SIC industry, we are able to get a comprehensive picture of

diversification patterns in the U.S. economy. Comparison with the Census of Manufacturers

indicates that Trinet contains 95 percent of all establishments it should (Voight, 1993), and

that omissions are most likely for small firms (which are less likely to be diversified). The

primary measure of SR_{ij} was calculated from the Trinet files of 1981, using all recorded firms

active in two or more 4-digit SIC-codes as a basis. After deleting single business firms,

government owned and non-profit industries, this resulted in a total of 13,164 diversified

firms, active in 929 different industries, covering 57,647 individual businesses. Of the

---

[14] Trinet also includes foreign establishments in the U.S.

[15] Because of changes in the parent coding in the Trinet database in 1979, and changes in the SIC classification scheme in 1987, only data from the years 1981, 1983 and 1985 are directly comparable.

431,056 possible industry pairs, 122,105 were observed. The measure of SR_{ij} between the observed industry pairs ranged from -7.97 to 93.55 with a mean of 4.33 and a standard deviation of 5.06. Based on these calculations of SR_{ij} we calculated WAR_i and $WARN_i$ following the procedures described in section 3. Finally, note that relatedness between industries as measured by SR_{ij} changes very little over the period covered in this study. The correlation between SR_{ij} in 1981 and 1983 is 0.941, and between 1981 and 1985 is 0.895.

4.2. Sample

H1 stated that if the survivor principle holds, businesses with low levels of survivor-based relatedness to *all other* businesses in a corporate portfolio (WAR_i) are more likely to be exited than businesses with high levels of survivor-based relatedness. H2 stated that businesses with low levels of survivor-based relatedness ($WARN_i$) to the closest *neighboring* businesses in the corporate portfolio are more likely to be exited than businesses with high levels of survivor-based relatedness. To construct our sample for examining these hypotheses, we began with all 13,164 diversified firms in the Trinet database. To obtain the necessary data for the variables of interest, Trinet data had to be merged with financial data from Compustat database. Since the parent identity numbers in these two databases are different, the matching had to be done alphanumerically by parent name. Spelling differences between the two databases resulted in undisputable matches for 854 companies that had entries in all the years needed to compute the variables. We believe this matching procedure to be a functional equivalent of random sampling, since there is no reason to expect spelling matches of parent names in the two databases to be biased in any particular way.[16]

---

[16] The matching of Trinet and Compustat data creates a bias toward larger firms compared to Trinet data alone. The reason for this is that Compustat contains publicly traded firms only, while Trinet contains both. Publicly traded firms are on average larger than privately held firms.

We imposed two further restrictions on our sample. One was to remove firms that were sold or liquidated in their entirety between 1981 and 1985. The reason being that such actions do not reveal information about the merits of combining different industries inside one firm, while exiting some businesses and keeping others do. Finally, we restricted our sample to firms that had more than 20 million dollars in sales in 1981. This resulted in a net sample of 70 firms. These 70 firms operated 2,640 businesses in 466 different 4-digit SIC codes in 1981. They exited a total of 593 industries between[17] 1981 and 1985, while they remained in 2,024 industries throughout the period and entered 738 new industries.

To test our hypotheses we included all the 593 instances of exit, but rather than using the entire sample of non-exits, we used a random-sample generator to select a sample of non-exits of comparable size. State based sampling has been suggested as preferable to a pure random sample when a population is overwhelmingly characterized by one state, and will provide unbiased and consistent coefficients for all variables except the constant term (McFadden and Manski, 1981). Thus the final sample consisted of a total of 1,191 observations, 593 of which were exits, and 598 non-exits.

To test H1 and H2 we developed a model of the relationship between the probability of exit and survivor-based relatedness, which controls for a number of industry- and parent variables that previous research indicates may affect the exit decision. Given the dichotomous nature of the dependent variable, a logistic regression analysis was considered appropriate for testing this model. The general model is the following:

$$P(\text{exit}=1) = \beta_1 + \beta_2 \text{(industry growth)} + \beta_3 \text{(industry concentration)} + \beta_4$$

$$\text{(industry profitability)} + \beta_5 \text{(parent size)} + \beta_6 \text{(parent market share)} + \beta_7$$

$$\text{(parent leverage)} + \beta_8 \text{(parent liquidity)} + \beta_9 \text{(parent relatedness)} + \varepsilon.$$

---

[17] Note that the relatedness measures are based on data from the 1981 files, while exits are identified using the 1983 and 1985 files (i.e., exits after 1981).

The dependent variable is coded as follows. If a parent active in a 4-digit SIC code in 1981 has exited this business by 1985, the dependent variable is given a value of 1. If the parent is still active in the industry, the value assigned is 0. Both divestures and closures are thus considered to represent exit. The Trinet database was used to identify exits and non-exits.

4.3. Industry-level independent variables

In testing H1 and H2 it is important to control for influences on the exit decision that are attributable to the profitability of the industry in question. All things equal firms are presumably less likely to exit an industry with high average profitability than one with low profitability. The negative relationship between relatedness and the probability of exit should exist independent of such industry effects. This study controls for three industry level variables that both theory and empirical research have found to affect the attractiveness of an industry. These variables are: industry growth, industry concentration, and industry profitability. Note that the latter variable, industry profitability, is included to control for unspecified industry effects not captured by the two other industry level control variables.

Industry growth. Industry growth is widely assumed to affect industry attractiveness favorably, because it allows firms to grow without having to steal customers from competitors. Thus, industry growth tends to soften competitive rivalry and raise the average profitability. Such a relationship has been confirmed in numerous empirical studies (i.e. Kwoka and Ravenscraft, 1986; Salinger, 1984; Schmalensee, 1989). One would accordingly expect a negative relationship between the growth of an industry, and the probability of exit. This variable is derived by estimating the growth in percent of industry sales between 1981 and 1985, as reported in Trinet.

Industry concentration. Traditional industrial organization theory posits a positive

relationship between industry concentration and industry profitability (Bain, 1956; Porter,

1980). Scale economies and other sources of market power, it is argued, reduce the threat

from potential entrants, allowing incumbents more room to raise prices without inviting

entry. Such a relationship has found support in empirical studies (Bain, 1951; Montgomery,

1985; Weiss, 1974), but the relationship between concentration and industry profitability

remains controversial since a number of studies have failed to find such a relationship, and

for those that do, the direction of causality is uncertain (see Schmalensee, 1989, for a review).

We expect a negative relationship between the industry concentration and the probability of

an industry being exited. The variable is derived by estimating the 4-firm concentration ratio

of each industry for 1981, based on the Trinet data.

Industry profitability. Industry profitability may be affected by numerous other factors

beyond growth and concentration. To control for such unspecified factors we calculated a

measure of the median return on assets for each industry over the 1980–82 period. The

procedure used here calls for some elaboration. The Compustat database which was used to

derive industry profitability consists of a segment database that report return on assets (ROA)

by 4-digit segment SIC code, and a corporate database that reports report ROA on the firm

level. We used all observations in the segment database and all single business firms in the

corporate database to calculate industry profitability. However this creates a problem because

the ROA measures are not directly comparable over these two databases. Because of

incomplete asset allocation, ROA is systematically higher in the segment database. To

preserve observations we calculated the mean of all observations in each database for each

year, and we subsequently divided each individual observation by this mean. Thus, the

individual observations were standardized as deviations in percent from the database mean

for the relevant year. This allowed us to use observations from both databases and all three

years, and we subsequently calculated the median of this measure for each industry. Where a minimum of five observations were obtained, this was done on the level of 4-digit SIC-industries, if less than five observations were obtained the same measure was calculated on the 3-digit level, if still less than five observations where obtained, the measure was calculated on the 2-digit level (following Berger and Ofek, 1995). We expect a negative relationship between industry profitability and the probability of exit.

4.4. Firm-level independent variables

Besides industry level factors, testing H1 and H2 also requires control for effects that are attributable to other properties of the firm than that of the relatedness of the business in question. We control for four firm level properties that both theory and empirical research have found to affect exit decisions. These variables are market share, parent size, parent leverage and parent liquidity.

Market share. A positive relationship between market share and profitability is documented in a large number of empirical studies (e.g., Gale, 1971; Sheperd, 1972; Robins and Wiersma, 1995). There are numerous explanations for this relationship, ranging from market power explanations through cost advantages due to learning curve effects and economies of scale. Given these positive performance effects we expect firms to be less likely to exit a business where they hold large market shares. The variable is measured as firm sales in industry *i* as percent of industry sales in 1981, and is expected to be signed negatively. The data are based on the Trinet files.

Parent size. The size of the parent has as noted been used as an indicator of market power and economies of scale. In addition parent size is an indicator of a parent's level of financial- and other resources. Based on this we expect a negative relationship between parent size and

46

the probability of exit. The variable is measured as the total sales of the parent in 1981, based on Trinet data.

Parent leverage. A highly leveraged parent may be under pressure from banks and investors to sustain a high cash flow in the short term to avoid an excessive bankruptcy risk. This is likely to reduce the patience with low performing businesses in the portfolio. Furthermore, a highly leveraged firm may experience constraints in funding attractive investment opportunities. Divesture of one business may therefore be an attractive way to finance investments in another. Therefore we expect the leverage of the parent to be positively related to the probability of a business being exited. The variable PARLEV is measured as long term debt to market value in 1981. The data were obtained from the Compustat database.

Parent liquidity. A low current ratio, like a high leverage ratio, may indicate financial constraints that makes exit more likely in order to reduce bankruptcy risk, or undertake divesture as a method of financing investments in other businesses. We therefore expect the current ratio to be negatively related to the probability of a business being exited. The variable is measured as the ratio of current assets to current liabilities in 1981. Data were obtained from the Compustat database.

4.4. Relatedness variables

The survivor-based relatedness measures employed in the test of H1 and H2 have been presented above. The measure used in testing H1 is called WAR_i and captures the sales weighted average relatedness of the business i to *all other* businesses in the parent k. The measure used in testing H2 is called $WARN_i$ and it captures the sales weighted average relatedness of business i to the two closest *neighboring* businesses in the parent k. Given that the survivor principle holds, we expect WAR_i and $WARN_i$ to be signed negatively.

Variable definitions, data sources, and predicted signs are summarized in Table 1 below, while Table 2 shows the means, standard deviations, and correlation coefficients for all independent variables.

[Tables 1 and 2 about here]

5. Results and discussion

The results from the logistic regression analyses are presented in table 3. Table 3 contains three different logistic regression models. Model 1 contains control variables only. Model 2 contains control variables plus the survivor-based measure of relatedness *WAR*, which is the sales weighted average relatedness of the industry i, to all other industries in the parent portfolio. Model 3 contains control variables plus the survivor-based measure of relatedness *WARN*, which is the sales weighted average relatedness of the industry i to the two closest neighboring industries in the parent portfolio.

[Table 3 about here]

H1 predicted that *WAR* would have a significant and negatively signed effect on the probability of exit. As can be seen from model 2 in table 3, this hypothesis is strongly supported. The coefficient of the variable *WAR* is indeed negative and highly significant. We furthermore note that all measures of model performance improve substantially in comparison with model 1. The model chi-square increases from 119.6 to 185.96 which is significant at the 0.001 level. The two pseudo-R^2 measures (Cox & Snell R^2 and Nagelkerke R^2) both increase more than 50 percent, and the ability to correctly predict exits also increases

approximately 50 percent. Table 4 shows how much explanatory power each variable adds to model 2, and we observe that WAR adds more than any other variable, including such "classics" as industry growth, industry concentration, industry profitability and market share.

[Table 4 about here]

H2 predicted that *WARN* would have a significant and negatively signed effect on the probability of exit. This hypothesis is also strongly supported. The coefficient is negative and highly significant. Again, all measures of model performance improve substantially in comparison with model 1. The model chi-square increases from 119.6 to 170.24 which is significant at the 0.001 level. The two pseudo-R^2 measures, Cox & Snell R^2 and Nagelkerke R^2, increase with 39 and 40 percent respectively. And the ability to correctly predict exits also increases by 50 percent. Table 5 shows how much explanatory power each variable adds to model 3, and again we observe that the survivor-based measure (in this case WARN) adds more than any other variable.

[Table 5 about here]

However, we also note that the coefficient of *WARN* is both slightly smaller and that the rate of improvement in the model performance over model 1 is slightly less for model 3 than for model 2. In our opinion this indicates that there are positive effects of relatedness that extends beyond the two closest related industries in a portfolio, and that the noted differences between model 2 and 3 reflect this. More specifically, *WARN* performs slightly below *WAR* because *WARN* restricts relatedness effects to the two closest related industries in each portfolio.

While H1 and H2 did receive strong support, some of the control variables did not behave as expected. This was particularly the case for industry concentration, which was expected to be negatively related to the probability of exit, but instead a significant positive relationship was found. Although surprising, this finding does have precedents (Schmalensee, 1989). One reason may be that concentrated industries are difficult to enter, and that some of the exits in our sample are in fact unsuccessful entry attempts. More generally, market concentration may be a poor measure of competition (Hayek, 1948). The conditions that cause concentration may also cause intense rivalry. For example the presence of substantial fixed costs will typically cause concentration, but it also tends to make battles for market share more intense. Thus episodes of intense competition may break out which hurt all incumbents, but smaller firms in particular. As a result these smaller firms may exit by closing down, divesture to a larger firm, or merger with other small firms to gain economies of scale. To test this we split the sample in two based on market share, which resulted in a significant positive relationship between concentration and probability of exit for the subsample with the smallest market shares, and no relationship for the subsample with the largest market shares.

Another surprising finding was the insignificant relationship between industry profitability and the probability of exit. There may be several factors contributing to this finding. One reason may be associated with what Goold et al. (1994) call the parenting advantage criterion. This criterion states that a business should be sold when the current owners are not the best possible owners, because the better owners will be willing to pay more than the current ones can expect from continued ownership. Abiding by this criterion would weaken the tendency to retain a business because it is profitable or in a profitable industry. Secondly, the distribution of profitability in high return industries may be highly skewed, such that even if the profitability of the median firm is high, it covers a tale of low profitability firms. Unfortunately the Compustat data used to construct the industry profitability measures in this

study are too coarse to explore this possibility further. Thirdly, our results may simply be an artifact of limitations in the underlying data used to compute industry profitability.

Finally, we may note that parent size, parent leverage and parent liquidity did not reach significance either. This is somewhat less surprising since the number of parents included in the final sample was 70. Thus we may lack sufficient statistical power to capture effects of the sizes associated with these variables.

6. Conclusions

In sum, both H1 and H2 received strong support. Apparently, in the context of diversification what firms combine frequently does *on average* represent more efficient combinations than those that occur rarely. In other words, what competitive firms do - does contain information about what is efficient. Thus it would seem that the comparative efficiency version of the survivor principle has withstood this first attempt at falsification, and more specifically, that relying on the survivor principle is defensible when testing hypotheses about the efficiency consequences of portfolio decisions.

However, these conclusions should be drawn with some caution. First of all there could be other explanations for the observed patterns in the data. Some would argue that the early 1980s was a period where refocusing and de-conglomeration was increasingly fashionable, and that the exits decisions observed here may be more influenced by fashion and herd behavior than efficiency.[18] To control for this possibility, future work should examine other time periods and use other efficiency measures than exit. Secondly, although the findings

---

[18] Note that *WAR* and *WARN* are both positively (and statistically significantly) correlated with market share, consistent with an efficiency explanation.

reported here does give support to the use of the survivor principle in research on diversification, this does not mean that it holds for other decisions, such as for example the choice between market- and hierarchical governance. To put the survivor principle on a firm empirical footing requires that the validity of the survivor principle is tested empirically in all areas where it is frequently used as an assumption. Obviously a substantial amount of work remains before this is accomplished, but as far as this study is concerned, we conclude that the survivor principle did survive diversification.

REFERENCES

Alchian AA. 1950. Uncertainty, Evolution, and Economic Theory. *Journal of Political Economy* LVIII: 211-221.

Anderson E. 1985. The Salesperson as Outside Agent or Employee: A Transaction Cost Analysis. *Marketing Science* 4(3): 234-254.

Bain JS. 1951. Relation of Profit Rate to Industry Concentration: American Manufacturing 1936-1940. *The Quarterly Journal of Economics* 65(3): 293-324.

Bain JS. 1956. *Barriers to new competition : their character and consequences in manufacturing industries*. Harvard University Press: Cambridge, Mass.

Barney JB. 1986. Strategic Factor Markets - Expectations, Luck, and Business Strategy. *Management Science* 32(10): 1230-1241.

Berger PG, Ofek E. 1995. Diversifications Effect on Firm Value. *Journal of Financial Economics* 37(1): 39-65.

Blaug M. 1980. *The Methodology of economics, or How Economists Explain*. Cambridge University Press: Cambridge.

Boerner CS, Macher JT. 2001. Transaction Cost Economics: An Assessment of Empirical Research in the Social Sciences. *Working Paper*: McDonough School of Business, Georgetown University.

Boland LA. 1979. Critique of Friedman Critics. *Journal of Economic Literature* 17(2): 503-522.

Caldwell BJ. 1980. A Critique of Friedman Methodological Instrumentalism. *Southern Economic Journal* 47(2): 366-374.

Dierickx I, Cool K. 1989. Asset Stock Accumulation and Sustainability of Competitive Advantage. *Management Science* 35(12): 1504-1511.

Dow GK. 1987. The Function of Authority in Transaction Cost Economics. *Journal of Economic Behavior & Organization* 8(1): 13-38.

Eisenhardt KM. 1985. Control - Organizational and Economic Approaches. *Management Science* 31(2): 134-149.

Elster J. 1989. *Nuts and Bolts for the Social Sciences*. Cambridge University Press: New York.

Enke S. 1951. On Maximizing Profits: A Distinction Between Chamberlin and Robinson. *American Economic Review* XLI: 566-578.

Farjoun M. 1994. Beyond Industry Boundaries - Human Expertise, Diversification and Resource-Related Industry Groups. *Organization Science* 5(2): 185-199.

Friedman M. 1953. The Methodology of Positive Economics. *Essays in Positive Economics*: 328. University of Chicago Press: Chicago.

Gale BT. 1971. Market Share and Rate of Return. *Review of Economics and Statistics* 54: 412-423.

Goold M, Campbell A, Alexander M. 1994. *Corporate-level strategy : creating value in the multibusiness company*. J. Wiley: New York.

Harrod R. 1938. Scope and Method of Economics. *The Economic Journal* 48(191): 383-412.

Hayek FAv. 1948. *Individualism and economic order*. University of Chicago press: London.

Hodgson G. 1993. *Economics and Evolution*. University of Chicago Press: Ann Arbor.

Hoskisson RE, Hitt MA. 1990. Antecedents and Performance Outcomes of Diversification - a Review and Critique of Theoretical Perspectives. *Journal of Management* 16(2): 461-509.

Kwoka JE, Ravenscraft DJ. 1986. Cooperation V Rivalry - Price-Cost Margins by Line of Business. *Economica* 53(211): 351-363.

Lester RA. 1946. Shortcomings of Marginal Analysis for Wage-Employment Problems. *The American Economic Review* 36(1): 63-82.

Matsusaka JG. 2001. Corporate diversification, value maximization, and organizational capabilities. *Journal of Business* 74(3): 409-431.

Mäki U. 1994. Reorienting the Assumptions Issue. In R Backhouse (Ed.), *New Directions in Economic Methodology*: x, 394. Routledge: London ; New York.

McFadden D, Manski CF. 1981. Alternative estimators and sample designs for discrete choice analysis. In D McFadden, CF Manski (Eds.), *Structural analysis of discrete data with econometric applications*: xxv, 477 s. MIT Press: Cambridge, Mass.

Montgomery CA. 1985. Product-Market Diversification and Market Power. *Academy of Management Journal* 28(4): 789-798.

Montgomery CA, Hariharan S. 1991. Diversified Expansion by Large Established Firms. *Journal of Economic Behavior & Organization* 15(1): 71-89.

Musgrave A. 1981. Unreal Assumptions in Economic-Theory - the F-Twist Untwisted. *Kyklos* 34(3): 377-387.

Nelson RR, Winter SG. 1982. *An Evolutionary Theory of Economic Change*. Belknap Press: Cambridge, Mass.

Penrose ET. 1952. Biological Analogies in the Theory of the Firm. *American Economic Review* XLII: 804-819.

Porter ME. 1980. *Competitive Strategy*. Free Press: New York.

Prahalad CK, Bettis RA. 1986. The Dominant Logic - a New Linkage between Diversity and Performance. *Strategic Management Journal* 7(6): 485-501.

Ramanujam V, Varadarajan P. 1989. Research on Corporate Diversification - a Synthesis. *Strategic Management Journal* 10(6): 523-551.

Robins J, Wiersema MF. 1995. A Resource-Based Approach to the Multibusiness Firm - Empirical-Analysis of Portfolio Interrelationships and Corporate Financial Performance. *Strategic Management Journal* 16(4): 277-299.

Salinger MA. 1984. Tobin-Q, Unionization, and the Concentration-Profits Relationship. *Rand Journal of Economics* 15(2): 159-170.

Schmalensee R. 1989. Inter-Industry Studies of Structure and Performance. In R Schmalensee, RD Willig (Eds.), *Handbook of Industrial Organization*. North Holland: New York.

Shelanski HA, Klein PG. 1995. Empirical-Research in Transaction Cost Economics - a Review and Assessment. *Journal of Law Economics & Organization* 11(2): 335-361.

Sheperd WG. 1972. The Elements of Market Structure. *Review of Economics and Statistics* 54: 25-37.

Silverman BS. 1999. Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. *Management Science* 45(8): 1109-1124.

Simon HA. 1947. *Administrative behavior; a study of decision-making processes in administrative organization*. Macmillan Co.: New York.

Stigler GJ. 1968. *The organization of industry*. R. D. Irwin: Homewood, Ill.

Teece DJ, Rumelt R, Dosi G, Winter S. 1994. Understanding Corporate Coherence - Theory and Evidence. *Journal of Economic Behavior & Organization* 23(1): 1-30.

Voight CW. 1993. The Trinet large establishment data. University of california at Los Angeles, CA: Anderson Graduate School of Management.

Weiss LV. 1974. The concentration-profits relationship and antitrust. In Goldschmidt et al. (Ed.). *Industrial Concentration: The New Learning*: 201-220. Little, Brown: Boston, MA.

Wernerfelt B. 1984. A Resource-Based View of the Firm. *Strategic Management Journal* **5**(2): 171-180.

Williamson OE. 1975. *Markets and hierarchies, analysis and antitrust implications : a study in the economics of internal organization*. Free Press: New York.

Williamson OE. 1985. *The economic institutions of capitalism : firms, markets, relational contracting*. Free Press: New York.

Williamson OE. 1988. Economics and Sociology. In G Farkas, P England (Eds.). *Industries, Firms, and Jobs*: 159-186. Plenum Press: New York.

Winter SG. 1964. Economic Natural-Selection and the Theory of the Firm. *Yale Economic Essays* **4**(1): 225-273.

Table 1: Variable Definitions, Data Sources and Predicted Signs

Variable	Definition	Data Source	Predicted Sign
Industry growth	Sales growth in industry i between 1981 and 1985	Trinet	-
Industry concentration	4-firm concentration ratio in 1981 in industry i	Trinet	-
Industry profitability	Industry median ROA 1980–82 in industry i	Compustat	-
Market share	Market share in industry i for parent k in 1981	Trinet	-
Parent size	Total sales of parent k in 1981	Trinet	-
Parent leverage	Leverage of parent k in 1981	Compustat	+
Parent liquidity	Current ratio of parent k in 1981	Compustat	-
WAR	Weighted average relatedness of industry i to all other industries in the portfolio of parent k	Trinet	-
WARN	Weighted average relatedness of industry i to the two closest neighboring industries in the portfolio of parent k	Trinet	-

Table 2 Independent Variables

	Variables	Mean	Standard deviation	Industry growth	Industry concentration	Industry profitability	Market share	Parent size	Parent leverage	Parent liquidity	WAR
	Industry variables										
1	Industry growth	0.335	0.605								
2	Industry concentration	27.24	18.08	0.16***							
3	Industry profitability	0.064	0.384	0.16***	-0.03						
	Firm variables										
4	Market share	2.22	5.49	0.14***	0.45***	-0.09***					
5	Parent size	63713	62907	-0.02	0.10***	-0.01	0.14***				
6	Parent leverage	0.179	0.094	0.03	-0.04	-0.07**	-0.08**	-0.25***			
7	Parent liquidity	1.64	0.49	-0.05	-0.08***	0.03	-0.03	-0.42***	-0.08***		
	Relatedness variables										
8	WAR	8.69	6.77	0.09***	0.09***	0.04	0.20***	-0.03	0.06**	0.00	
9	WARN	23.67	11.92	0.16***	0.13***	0.04	0.18***	0.04	0.03	-0.05*	0.56***

Means, standard deviations, and correlation coefficients for independent variables: ****, ***, **, and * indicate correlations significant at the 1, 5, and 10 percent level, respectively. N=1,191.

Table 3: Results

	Model			Signs	
	(1)	(2)	(3)	*Expected*	*Observed*
Constant	0.44	1.02***	1.33***		
	(1.60)	(7.69)	(12.25)		
Industry growth	–0.69***	–0.68***	–0.59***	–	–
	(30.61)	(28.09)	(21.66)		
Industry concentration	0.01*	0.01*	0.01*	–	+
	(3.18)	(3.13)	(3.45)		
Industry profitability	–0.06	–0.02	–0.07	–	
	(0.15)	(0.01)	(0.17)		
Market share	–0.15***	–0.12***	–0.12***	–	–
	(37.39)	(25.51)	(28.82)		
Parent size	0.00	0.00	0.00	–	
	(0.50)	(1.12)	(0.56)		
Parent leverage	–0.61	–0.33	–0.45	+	
	(0.78)	(0.21)	(0.41)		
Parent liquidity	0.01	0.04	–0.03	–	
	(0.00)	(0.09)	(0.52)		
WAR	——	–0.09***	——	–	–
		(53.63)			
WARN	——	——	–0.04***	–	–
			(46.80)		
Log likelihood	1,531.46	1,465.10	1,480.82		
Model χ^2	119.59***	185.96***	170.24***		
$\Delta \chi^2$ vs. model 1		66.37***	50.65***		
Cox and Snell R^2	0.096	0.145	0.133		
Nagelkerke R^2	0.127	0.193	0.178		
Percent correct predictions	60.5	65.4	65.7		

Logistic regressions of the probability of exit on relatedness (*WAR* and *WARN*) and industry and parent characteristics. Wald statistics in parentheses. ***, **, and * represent statistical significance at the 1, 5, and 10 percent levels, respectively. N=1,191.

58

Table 4: Model if term removed (model 2)

Variable	Change in -2 Log Likelihood	Signifcance of change
Industry growth	30.89	0.00
Industry concentration	3.15	0.08
Industry profitability	0.01	0.93
Market share	38.91	0.00
Parent size	1.12	0.29
Parent leverage	0.21	0.65
Parent liquidity	0.09	0.77
WAR	66.37	0.00

Table 5: Model if term removed (model 3)

Variable	Change in -2 Log Likelihood	Signifcance of change
Industry growth	23.96	0.00
Industry concentration	3.48	0.06
Industry profitability	0.17	0.68
Market share	45.79	0.00
Parent size	0.56	0.45
Parent leverage	0.41	0.52
Parent liquidity	0.05	0.82
WARN	50.65	0.00

Yet Another Way of Measuring Relatedness - This One: Let Competition Do It!

Lasse B. Lien

Institutt for strategi og ledelse

Norges Handelshøyskole

Breiviksveien 40

N-5045 Bergen

Norway

lasse.lien@nhh.no

Paper no. 2

Abstract

Over the last two decades numerous procedures for measuring relatedness have been suggested, primarily driven by the inability of our empirical research to demonstrate consistent support for the hypothesized relationship between relatedness and corporate performance. The main problem by most accounts is associated with measuring relatedness in a manner consistent with theoretical developments, which includes some conditions that are very difficult to operationalize. This paper suggests a fundamentally different approach to these challenges. We submit a test of *a survivor-based approach* to measuring relatedness - where a combination of the selection processes in competitive markets and the wisdom of local decision makers are substituted for the insight of the researcher or the SIC-system in capturing relatedness.

1. Introduction

Despite the fact that no single issue has been given more attention in research on corporate

level strategy than the possible link between relatedness and performance, the empirical

results are usually summarized as mixed or confusing (Hoskisson and Hitt, 1990; Markides

and Williamson, 1994, 1996; Reed and Luffman, 1986; Robins and Wiersema, 1995). Most

reviewers seem to ascribe this dim state of affairs to problems associated with measuring

relatedness in a manner consistent with the theoretical developments, which includes some

conditions that are very difficult to operationalize. Given these problems we suggest a

fundamentally different approach, one where a combination of the competitive process and

the knowledge of local decision makers replace the wisdom of the researcher (or the SIC-

system) in capturing relatedness. Basically this amounts to an assumption that what is related

is what firms in competitive markets combine often. More specifically it implies that a

measure of the relatedness between a pair of industries can be obtained by considering how

often a pair of industries are actually combined inside a firm - compared to what one would

expect if diversification patterns were random. Industries are related when this difference is

large and positive, and they are unrelated if it is negative. This line of thinking was originally

suggested by Teece, Rumelt, Dosi and Winter (1994), but these authors only used it to

illustrate that coherence (non-randomness) was a salient attribute of the diversification

patterns of US firms. In other words, they did not make any attempt to evaluate this procedure

as a way of capturing inter-industry relatedness, which is what the present paper does.

Such a survivor-based procedure represents an abdication in terms of pinpointing what

relatedness *is* in specific instances, but could be potentially useful for studying what

relatedness *does* - in terms of influencing other variables of interest (i.e. performance, entry

mode, financing decisions, organizational parameters, etc.). To investigate this survivor-based

approach to relatedness we compare two measures built on the conventional SIC-based

measurement procedures with parallel measures derived from the survivor-based logic in terms of explaining exit decisions by diversified firms. We submit that survivor-based measures may be a useful complement to other approaches suggested in the literature, both the conventional SIC-based measures and the more recent innovations (Farjoun, 1994; Markides and Williamson, 1996; Robins and Wiersema, 1995; Silverman, 1999).

We proceed as follows: Section 2 summarizes the theoretical arguments behind the relatedness hypothesis, and from this we derive the conditions necessary for the hypothesis to hold. Section 3 discusses the most frequently used measurement procedures in existing research, and point out that important theoretical conditions are not captured by these procedures. Section 4 presents an alternative procedure based on the survivor logic. Two variants of a survivor-based measure of relatedness are then developed, and hypotheses contrasting these with equivalent measures based on the SIC-system are formulated. Section 5 discusses methodological issues. Section 6 presents our empirical findings, and section 7 concludes.

2. The Relatedness Hypothesis: The Theoretical Foundations

The relatedness hypothesis in its most rudimentary form states that multi-business firms that are constructed with portfolios of businesses that are similar (related) will perform better than portfolios of heterogeneous businesses. This immediately raises two questions. What are the relevant kinds of similarity? And under which circumstances will such similarities provide advantages for a diversifier that cannot be replicated by a non-diversifier or an unrelated diversifier? The absolute minimum requirement for relevant similarity is that resources[19] in

---

[19] We do not distinguish between resources and competences here, and resources should be interpreted as encompassing both

one industry are substitutes or complements to resources in another. If neither of these conditions are satisfied (either statically or dynamically) the word *related* in related diversification seems without economic content. However, while they are necessary, they are not sufficient.

To see this, let us first start with the classic situation involving resources that are *substitutes across* industries (i.e. economies of scope). Consider a situation where a resource in industry A is a perfect substitute for a resource in industry B. Under the standard microeconomic assumption of perfectly divisible resources, this perfect substitutability does not provide any advantage to a related diversifier active in both A and B.[20] In contrast, if the condition of perfect divisibility[21] is relaxed, a potential advantage may exist because a single business firm or an unrelated diversifier would be left with some costly excess capacity (Willig. 1979). Penrose (1959) is usually regarded as the first to relax the assumption of perfect divisibility. In her acclaimed account of the growth of firms she pointed out that excess capacity arises both because some resources are inherently indivisible (i.e. half a truck) but more importantly because firms in the course of their normal operations, as a result of learning, continuously generate new resources - and excess capacity in existing resources. However, unless there is some disadvantage associated with costly excess capacity for firms that do not pursue related diversification, no performance differential can be expected. Accordingly, we must add the condition that some indivisibilities exist.

However, Teece (1980, 1982) made the point that while the existence of indivisibilities explains joint production, it does not explain why joint production must be organized within a single firm. If the excess capacity created by indivisibilities can be traded in well functioning markets, single business firms and unrelated diversifiers will face no disadvantage. They can

---

[20] Disregarding for the moment issues of complementarity

[21] In this we include as a special case the situation where the resource in question is a public input, so that excess capacity will always exist.

simply sell or rent out their excess capacity, or alternatively, buy exactly the amount of capacity they need from others. In such a situation the related diversifier may actually compete at a disadvantage relative to single business firms, since related diversification will normally involve some measure of weakened incentives, increased complexity, and added overhead. However, the presence of some form of market failure can neutralize these concerns, because missing markets or high transaction costs (Teece, 1980, 1982; Williamson, 1985) may make market contracting unfeasible.

Summing up, even if resources are perfect substitutes across industry boundaries, there is no reason to expect the relatedness hypothesis to hold, unless there also exists some indivisibilities that create an excess capacity problem for those not properly diversified. But even when this is the case, unless trading in this excess capacity is subject to some form of market failure, we should not expect the relatedness hypothesis to hold.

More recently, several authors have suggested a shift in focus from resource substitutability to resource complementarity (Teecce et. al, 1994; Christensen and Foss, 1997; Foss and Christensen, 2001). Complementarity refers to situations where there are positive externalities (which essentially is a different form of indivisibility) across industry boundaries, either because the value of resources in one industry is positively affected by investments in resources in another, or because decisions about how resources are used in one industry affects how they should optimally be used in another. These positive spillovers create a quantitative and qualitative coordination problem which may bestow an advantage on a diversified firm (Richardson, 1972; Milgrom and Roberts 1992). This, however, requires that that a diversified firm can solve the coordination problem in ways that a single business firm or an unrelated diversifier cannot, which in turn requires that some form of market failure

exists. Hence. the condition of market failure also applies in situations involving complements.

Of particular importance to these authors are *dynamic* complementarities, which refers to an ability to exploit *diversity* (i.e. resources residing in different businesses) to identify new ways of combining existing resources, or speed up the development of new resources. The benefits to *similarity* in this context would seem to arise because the magnitude of such dynamic complementarities may be larger if the industries in question share some basic features (March. 1991), and also because some commonalities may facilitate their exploitation (Prahalad and Bettis. 1985). The amount of dynamic complementarity between industries would therefore seem to depend on the balance between variety and similarity (Christensen and Foss. 1997). Industries with the right balance between variety and similarity would be expected to produce larger dynamic complementarities than industries that are either too different or too similar.[22]

The implications for the relatedness hypothesis are that a portfolio of businesses with strong inter-industry complementarities should be considered related (or coherent), and should ceteris paribus outperform unrelated portfolios and single business firms. Given. of course, that such firms cannot create and exploit complementarities equally well by means of market contracting.

In sum. any kind of similarity between industries cannot be expected to bring about an advantage to related diversifiers. Indivisibilities (either in the form of productive capacity or positive externalities) must exist, and market contracting must not leave an unrelated

---

[22] It is pertinent to note that for instance Christensen and Foss (1997) deny that it is meaningful to discuss relatedness between industries, since in their view relatedness is specific to the individual firm. We disagree with this position, because we believe that there are important resource commonalities within industries, which makes it meaningful to discuss relatedness between industries. However, we acknowledge that since there are also resource differences inside industries, there is a firm specific component in relatedness. However, consistent with most authors discussing the topic of relatedness, we focus on the inter-industry component of relatedness.

diversifier or a single business firm equally well off. Unfortunately, to develop measures of relatedness that screen effectively for these conditions in a convincing manner is extremely challenging, and as we discuss in the next section, existing measures seem mainly to tap into the degree to which resources in one industry can function as substitutes for resources in another.

3. Existing Measures of Relatedness

Measures of relatedness within the existing strategic management research can be divided into three major categories: categorical measures, continous SIC-based measures, and recent developments. As argued below there are problems associated with all of these measures, in particular with their ability to capture all the conditions specified in the preceding paragraph.

Categorical Measures

The categorical approach is dominated by the work of Rumelt (1974) which has become the standard for the use of categorical measures. Based on three ratios, Rumelt classified diversification strategies into four broad categories (nine if subcategories are included). These were: single business firms, dominant business firms, related firms, and unrelated firms. The ratios used for classification were:

- *Specialization ratio:* The proportion of a firm's revenue that can be attributed to its largest single business
- *Related ratio:* The proportion of a firm's revenue that can be attributed to its largest group of related businesses
- *Vertical ratio:* The proportion of a firm's revenue that arise from all byproducts, intermediate products, and end products of a vertically integrated sequence of processing activities

An important element is the definition of what constitutes a related business, which subsequently affects the related ratio. This was done subjectively using similarities in inputs,

production technology, distribution channels and customers. As many others have pointed

out, there are potential problems with this procedure. One is of course the reliability- and

intersubjectivity problems associated with the subjective element in the classifications,

another is that by measuring relatedness on a nominal level it only allows for comparisons of

within group averages and the procedure therefore becomes quite restrictive.

These concerns not withstanding, an equally important problem is that the procedure mainly

seems to capture the degree to which one can assume that resources are potential substitutes

across industry boundaries. This will probably be the case when there are "large similarities in

inputs, production technology, distribution channels and customers". The measure clearly

does not capture whether there are indivisibilities present for the substitutable resources, and

hence whether excess capacity is likely to develop, nor does it contain any notion of market

failure that makes contracting over the relevant resources costly. Both of these conditions

were above identified as necessary conditions for economies of scope to create an advantage

for the related diversifier, hence such a measure would be prone to exaggerate relatedness in

some instances (i.e. where resources are close substitutes, but these additional conditions are

not met). The implicit focus on similarities and economies of scope also raises the concern

that such a procedure may not capture (dynamic) complementarity well, which implies that it

will underestimate relatedness in other instances (Foss and Christensen, 2001).

Given these shortcomings vis a vis the theoretical conditions, categorical measures have

important limitations in terms of testing the relatedness-performance relationship, unless these

conditions are somehow otherwise controlled for.

Continous SIC-based Measures

The continous SIC-based measures are currently the most widely used approach (Robins and

Wiersema, 2003). They include measures such as the entropy index (Jaquemin and Berry,

1979), and the concentric index (Caves, Porter and Spence, 1980).[23] The advantage of using

measures based on the SIC-system is that there is no subjective element in classifying the

degree of relatedness, and it also allows relatedness to be measured on an interval level. The

2, 3 and 4-digit levels in the SIC-system are treated as points on an underlying scale of

relatedness, and arithmetic values are assigned to the distances. This allows for a wide range

of statistical operations, and use of the large amounts of secondary data available in the SIC-

format.

However, the use of distances in the SIC-system also introduces problems, in that it imposes

some very strong assumptions on the SIC-system. It assumes that industries are homogenous

within category levels, which is problematic if the breadth of the industry classifications vary.

In fact most observers agree that they do (Robins and Wiersema, 1995: Rumelt 1982). The

second assumption is that it assumes that there is equal dissimilarity between real industries

when they are equally distant within the SIC-hierarchy. This assumption is also problematic.

Our major point, however, is that SIC-based measurements do not fare any better than the

categorical measures in terms of capturing the conditions of indivisibility and market

failure.[24] This means that even if we believe that there is a high correlation between distances

in the SIC-system and the degree to which resources are functional substitutes across

---

[23] The concentric index is calculated as follows :

$$FDIVERS_k = \Sigma_i P_{ki} \; \Sigma_j \; P_{kj} d_{ij}$$

Where:

P_{ki} = percentage of sales for firm k in industry i

P_{kj} = percentage of sales for firm k in industry j

d_{ij} = weighting factor such that $d_{ij} = 0$ where i and j belong to the same 3-digit SIC-category, $d_{ij} = 1$ where i and j belong to the same 2-digit category but different 3-digit categories, and $d_{ij} = 2$ where i and j are in different 2-digit categories.

The related portion of entropy is calculated as follows:

$$E_R = E_T - E_U = \Sigma P_T \; ln(1/P_T) - \Sigma P_U \; ln(1/PU)$$

Where:

E_R = Related component of entropy

E_T = Entropy defined at the 4-digit level

E_U = Entropy defined at the 2-digit level

P_T = Percentage of sales in each 4-digit industry

P_U = Percentage of sales in each 2-digit industry

[24] A study that does consider the condition of excess capacity is Chatterjee and Wernerfelt (1991).

industries, the omission of these two conditions will imply that this kind of relatedness is prone to be exaggerated. And according to Foss and Christensen (2001) the SIC-based procedure has an implicit bias towards economies of scope, which indicates that it is not likely to capture dynamic complementarities well, suggesting that this type of relatedness is prone to be underestimated. The combined effect of over- and underestimation of relatedness is that it becomes problematic to test the relatedness-performance relationship using continuous SIC-based measures.

Recent Developments

One possible avenue for closing the gap between theoretical conditions and measurement would be to focus on categories of resources that are more likely than others to generate excess capacity and positive externalities (i.e. imperfect divisibility), and be subject to market failure. If such categories can be identified, similarities with respect to such resources should be more likely to have a positive effect on the relative performance of a related diversifier. One recent approach seems particularly promising in this respect. We are here referring to the studies that use patent filings as an indication of technology flows between industries, which in turn indicate to what extent technological resources in one industry are valuable in another (Laursen and Meliciani, 2000; Robins and Wiersema, 1995; Silverman, 1999[25]; Piscitello, 2000). These studies seem to conform more closely to theoretical predictions than the average of the studies based on SIC-data and categorical measures, a finding which is possibly explained by a better match with theoretical conditions. There are indeed reasons to believe that technological resources are often imperfectly divisible, because several types of technological resources -patents in particular- can be described as quasi-public goods. This means that the use of such resources in one business does not preclude its use in another (if

---

[25] A particularly interesting study is Silverman, 1999 which explicitly examines the influence of market failure on patterns of diversification

the technology or knowledge can be replicated at zero or low marginal cost). It is also

plausible that such technology flows are a reasonably good indicator of the extent to which

there are dynamic complementarities across industries. Finally, there are reasons to believe

that technological resources are often subject to market failure problems that make them

costly to trade (Teece, 1986). Some technological resources may have a tacit element which

makes contracts related to their transfer difficult to enforce. Alternatively, technological

resources may be easy to transfer, but the value they represent may be difficult to appropriate

once the content of the knowledge has been revealed to a potential buyer.

However, the procedure is not without limitations. First of all the measures using patent data

can only capture relatedness associated with technological resources, and do so in industries

where patenting is a non negligible phenomenon. This indicates some severe restrictions on

where these measures can be applied, and the forms of relatedness they can capture. In

addition these measures are probably noisy even under favorable circumstances. Not all

technological resources are quasi-public goods, for example if they are somehow linked to

knowledge that cannot be easily replicated. This can be the case if technology transfer

requires knowledge that cannot be separated from one or a group of individuals (due to for

example tacitness). The capacity of these individuals may be exhausted in existing

applications, leaving the condition of excess capacity unsatisfied. Also, there are several types

of technological resources where market trading is indeed feasible (Levin et al., 1987; Teece,

1986; Silverman, 1999). The phenomenon of licensing arrangements for technology and

patent trading illustrates this point.

In addition to the alternatives already presented, there is a plethora of less frequently used

measures. These include the use of human resource profiles (Farjoun, 1994), input ratios

(Montgomery and Hariharan, 1991), commodity flows (Fan and Lang, 2000), and more.

While these may vary in their ability to capture to what extent resources in one industry are

substitutes or complements to resources in another. none of them offer a full solution for capturing the conditions of indivisibility and market failure. In sum. all known alternatives for measuring relatedness are flawed in terms of representing the theoretical developments with respect to the relatedness hypothesis. The survivor-based approach we are about to suggest is also less than perfect. but the question is whether it is less so than the alternatives.

4. A Survivor-based Approach to Relatedness

What has become known as the survivor principle (henceforth: SP) has as it core idea the notion that the competitive process screens for efficiency - and does so well enough that a sample of competitive firms will be dominated by the decisions or behaviors that are efficient - at least in a comparative sense (Alchian. 1950: 211).[26] Two key processes are given the burden of ensuring this. One is that firms making negative profits will. unless some corrective measure is taken. loose resources and ultimately become extinct. while firms making positive profits will acquire resources and grow. The other is that the desire to make positive profits provides a strong incentive for the less successful firms to imitate the more successful firms. While few believe that the competitive process performs this screening perfectly[27] the behavior of researchers in the field of economics. organizational economics and strategic management indicate quite an optimistic view of this process. After all theories or hypotheses about what is efficient are routinely tested by measuring what firms actually do. which indicates a belief in the basic conjecture of the SP. Examples include empirical tests of transaction cost analysis. where for example the hypothesis that vertical integration is more efficient than market governance when asset specificity is high. is tested by measuring whether firms actually integrate when assert specificity is high (e.g. Joskow, 1985;

---

[26] For the view that the competitive process creates outcomes that are optimizing, cfr. Friedman (1953).
[27] Milton Friedman is a possible exception (Friedman, 1953).

Monteverde and Teece, 1982). We also recognize it from agency theory where hypotheses about the relative efficiency of alternative contracts are tested by measuring which contracts firms actually employ (e.g. Anderson, 1985; Eisenhardt, 1985). And we recognize it from studies of diversification within the strategic management literature, where for example hypotheses about what constitutes efficient patterns of diversification are tested by measuring their consistency with actual patterns of diversification (e.g. Farjoun, 1994; Montgomery and Hariharan, 1991; Silverman, 1999).

If the basic conjecture of the survivor principle is valid, a measure of relatedness can be built on the idea that what is related is what firms in competitive markets combine often. The fundamental premise of this survivor-based approach to relatedness is thus that industries that are related will be more frequently combined within a firm. More specifically, we estimate how much the frequencies of *actual* combinations of four-digit SIC industries deviate from what one would expect if diversification patterns were *random*. We take this difference to constitute a survivor-based measure of the relatedness between a pair of industries.

A survivor-based measure of relatedness has the potential advantage that it incorporates the knowledge of the best informed actors (which presumably are those making portfolio decisions), but even if their information is poor, their decisions have been screened by the competitive process, which will enforce a reversal of poor decisions. Therefore it is not implausible that a survivor-based measure is better at capturing relatedness than the existing alternatives. On the other hand, the behavior of decision makers is surely not optimal, and screening function of the competitive process is surely not perfect either. A survivor-based measure will therefore include noise. Several authors criticizing the SP, have indicated that they believe the level of noise will be substantial (e.g. Elster, 1989; Hodgson, 1993; Winter 1971). However, in terms of measuring relatedness and testing the relatedness hypothesis the existing alternatives involve substantial noise too. The question is therefore one of relative

noise, and the best judge on this issue is data. We now move on to describe in detail how a survivor-based measure of relatedness can be constructed. Our approach is based on a procedure originally developed by Teece, Rumelt, Dosi and Winter (1994).

Let the universe of diversified firms consist of K firms, each active in two or more of I industries. Let $C_{ik} = 1$ if firm k is active in industry i. The number of industries participated in by firm k is $m_k = \Sigma_i C_{ik}$ and the number of diversified firms present in industry i is $n_i = \Sigma_k C_{ik}$. Let J_{ij} be the number of diversified firms active in both industries i and j, such that

$J_{ij} = \Sigma_k C_{ik} C_{jk}$. Thus J_{ij} is a count of how often industries i and j are actually combined within the same firm. J_{ij} will be larger if industries i and j are related, but will also increase with n_i and n_j. To remove the effect of the size of industries i and j, the number J_{ij} is compared with the number of expected combinations if diversification patterns were random.

The random diversification hypothesis can be operationalized as a hypergeometric situation where a sample of size n_i is drawn (without replacement) from a population of K firms. Those chosen are considered active in industry i. A second independent sample of size n_j is then drawn from the population of K firms. Those chosen are considered active in industry j. The number x_{ij} of firms active in both i and j is then a hypergeometric random variable with population K, special members n_i and sample size n_j. The distribution function for this variable is then:

$$\Pr(X_{ij} = x) = f_{hg}(x, K, n_i, n_j) = \frac{\binom{n_i}{x}\binom{K - n_i}{n_j - x}}{\binom{K}{n_j}}$$

The mean and variance of X_{ij} are:

$$\mu_{ij} = E(X_{ij}) = \frac{n_i n_j}{K}.$$

$$\sigma^2 = \mu_{ij}\left(1 - \frac{n_i}{K}\right)\left(\frac{K}{K-1}\right).$$

A standardized measure of the relatedness between industries i and j is then constructed based on the difference between J_{ij} and μ_{ij} in the following fashion:

$$SR_{ij} = \frac{J_{ij} - \mu_{ij}}{\sigma_{ij}}$$

The measure SR_{ij} is thus a standardized measure of how much actual combinations exceed expected combinations under the random diversification hypothesis. With this fundamental measure of the relatedness between a pair of businesses it is possible to compute various relatedness measures.

The two variants chosen here (elaborated below) reflect our choice of dependent variable, which is the probability that a given business will be exited. The reason being that exit is a performance measure not subject to manipulation or random variation, while for example ROA is subject to both. Exit by closure is readily interpretable as a sign of performance below expectations, while exit by divesture is an admittance of lower expected performance than the acquirer. We expect a negative correlation between relatedness and the probability of exit for two reasons. One is our belief that relatedness will have a positive impact on performance (when the conditions specified above are met), the other is that we expect related businesses to be more closely integrated with other businesses, making them more difficult to divest. In comparing different measures of relatedness, our approach will be to compare their relative ability to explain the probability that a given parent will exit a given business.

The first measure captures the weighted average relatedness of a business i to all other businesses in the parent portfolio. Assume a diversified firm that participates in m industries. Its business in industry i has sales of s_i and survivor-based relatedness SR_{ij} with industry j. The weighted average relatedness $SURVTOT_i$ of the business in industry i to all other business in the firm is then defined as:

$$SURVTOT_i = \frac{\sum_{j \neq i} SR_{ij} s_j}{\sum_{j \neq i} s_j}$$

A parallel measure based on SIC-distances can be obtained as follows:

$$SICTOT_i = \frac{\sum_{j \neq i} d_{ij} s_j}{\sum_{j \neq i} s_j}$$

Where $d_{ij} = 2$ if i and j are in the same 3-digit SIC codes

$d_{ij} = 1$ if i and j are in different 3-digit-, but the same 2 digit SIC codes

$d_{ij} = 0$ if i and j are in different 2-digit SIC codes

Note that this approach builds on the concentric index (Caves et al., 1980), but has been modified to accommodate our choice of dependent variable (a precedent of this modified concentric index is Sharma, 1998).

An alternative approach does not consider how related each business is to *all* other business in the corporate portfolio, but how related each business is to the parent's core business (measured as the largest 4-digit SIC code). Let SR_{ic} be the survivor-based measure between a

given industry i, and the core business c. In this measure we include consideration of the size of the focal business (where size refers to share of the parent's total sales), and construct it so that its value is high for large businesses close to the core business, and smaller for a small businesses distant to the core business. This measure $SURVCORE_i$ is then defined as:

$$SURVCORE_i = SR_{ic} \frac{S_i}{\sum_i S_i}$$

Again we also computed a parallel measure based on SIC-distances:

$$SICCORE_i = d_{ic} \frac{S_i}{\sum_i S_i}$$

Where $d_{ic} = 3$ if $i = c$

$d_{ic} = 2$ if $i \neq c$ and i and c are in the same 3-digit SIC codes

$d_{ic} = 1$ if i and j are in different 3-digit SIC-codes, but similar 2-digit SIC codes

$d_{ic} = 0$ if i and c are in different 2-digit SIC codes

Given these four measures: $SURVTOT_i$, $SICTOT_i$, $SURVCORE_i$, and $SICCORE_i$, we can formulate the following hypotheses.

H1: $SURVTOT_i$ will explain the probability exit significantly better than $SICTOT_i$

H2: $SURVCORE_i$ will explain the probability exit significantly better than $SICCORE_i$

H3: SURVTOT$_i$ and SURVCORE$_i$ will both explain the probability of exit significantly

better than SICTOT$_i$ and SICCORE$_i$

Note that while H1 and H2 states that survivor-based measures of relatedness will perform

better than parallel measured based on SIC-data. H3 states that both survivor-based measures

will perform better than both SIC-based measures. Note also our expectation that all four

measures will be negatively signed.

5. Methodology

This study involved two distinct empirical operations. First we had to calculate the

fundamental survivor-based measure of relatedness SR_{ij} for all pairs of industries in the US

economy. With this fundamental measure in hand we were able to calculate the survivor-

based measures SURVTOT$_i$ and SURVCORE$_i$ for any specific business belonging to any

specific parent. The second empirical operation was to conduct a test of our hypotheses.

linking these two measures and their SIC-based equivalents to the probability of exit.

5.1 Calculating SR_{ij}

To calculate SR_{ij} we used the AGSM/Trinet Large Establishment Database (Trinet). The

Trinet database contains records of all US establishments[28] with more than 20 employees.

including variables such as four-digit SIC code. corporate ownership and sales. By

aggregating the establishments for each parent in each four digit sic-code. and the different

four digit sic-codes for each parent. and different parents for each four-digit SIC industry, we

are able to get a comprehensive picture of diversification patterns in the US-economy.

---

[28] Trinet also includes foreign establishments in the US

Comparison with the Census of Manufacturers indicate that Trinet contains 95% of all establishments it should (Voight, 1993), and that omissions are most likely for small firms (which are less likely to be diversified). The primary measure of SR_{ij} was calculated from the Trinet files of 1981, using all recorded firms active in two or more four-digit SIC-codes as a basis. After deleting single business firms, government owned and non-profit industries, this resulted in a total of 13.164 diversified firms, active in 929 different industries, covering a total of 57.647 individual businesses. Of the 431.056 possible industry pairs, 122.105 were observed. The measure of SR_{ij} between the observed industry pairs ranged from −7.97 to 93.55 with a mean of 4.33 and a standard deviation of 5.06. Based on these calculations of SR_{ij}, we calculated measures of $SURVTOT_i$ and $SURVCORE_i$ by following the procedures described in section 4.

4.2 Testing H1-H3

Sample

The sample for testing H1-H3 was derived as follows. We started out with all the 13164 diversified firms in the Trinet database. To obtain the necessary data for the variables of interest, Trinet data had to be merged with financial data from Compustat database. Since the parent identity numbers in these two databases are different, the matching had to be done alphanumerically by parent name. Spelling differences between the two databases resulted in undisputable matches for 854 companies that had entries in all the years needed to compute the variables. We believe this matching procedure to be a functional equivalent of random sampling, since there is no reason to expect spelling matches of parent names in the two databases to be biased in any particular way.[29]

---

[29] The matching of Trinet and Compustat data creates a bias toward larger firms compared to Trinet data alone. The reason for this is that Compustat contains publicly traded firms only, while Trinet contains both. Publicly traded firms are on average larger than privately held firms.

We imposed two further restrictions on our sample. One was to remove firms that where sold or liquidated in their entirety between 1981 and 1985. The reason being that such actions do not reveal information about the merits of combining different industries inside one firm, while exiting some businesses and keeping others do. Finally, we restricted our sample to firms that had more than 20 million dollars in sales in 1981. This resulted in a net sample of 70 firms. These 70 firms operated 2640 businesses in 466 different four-digit SIC codes in 1981. They exited a total of 593 industries between[30] 1981 and 1985, while they remained in 2024 industries throughout the period, and entered 738 industries.

To test our hypotheses we included all the 593 instances of exit, but rather than using the entire sample of non-exits, we used the random sample generator in SPSS to select a sample of non-exits of comparable size. State based sampling has been suggested as preferable to a pure random sample when a population is overwhelmingly characterized by one state, and will provide unbiased and consistent coefficients for all variables except the constant term (McFadden and Manski, 1981). Thus the final sample consisted of a total of 1191 observations, 593 of which were exits, and 598 non-exits.

Statistical Methods

In order to test H1-H3 we developed a model of the relationship between the probability of exit and relatedness, which controls for a number of industry- and parent variables that previous research indicates may affect the exit decision. Our primary reason for including these control variables is to reduce the risk that the performance of one or more of the presented relatedness measures are inflated or deflated because of associations with such "other" factors, but in addition, we include control variables because we are interested in examining whether our data support the basic relatedness hypothesis. Given the dichotomous

---

[30] Note that we are focusing on businesses a parent operated in 1981, which may or may not have been exited by the year 1985.

nature of the dependent variable, a logistic regression analysis was considered appropriate for

testing our model. The general model is the following:

$P(\text{Exit}=1) = \beta_1 + \beta_2(\text{Industry growth}) + \beta_3(\text{Industry concentration}) + \beta_4(\text{Industry profitability})$

$+ \beta_5(\text{Parent size}) + \beta_6(\text{Parent market share}) + \beta_7(\text{Parent leverage}) + \beta_8(\text{Parent liquidity}) +$

$\beta_9(\text{Parent relatedness}) + \varepsilon$

Industry Level Independent Variables

One possible influence on the exit decision is the profitability of the industry in question. All

things equal firms are presumably less likely to exit an industry with high average

profitability than one with low profitability. A negative relationship between relatedness and

the probability of exit should exist independent of such industry effects. This study controls

for three industry level variables that both theory and empirical research have found to affect

the attractiveness of an industry. These variables are: industry growth, industry concentration,

and industry profitability. Note that the latter variable, industry profitability, is included to

control for unspecified industry effects not captured by the two other industry level control

variables.

INDGRSAL. Industry growth is widely assumed to affect industry attractiveness favorably,

because it allows firms to grow without having to steal customers from competitors. Thus,

industry growth tends to soften competitive rivalry and raise the average profitability. Such a

relationship has been confirmed in numerous empirical studies (i.e. Kwoka and Ravenscraft,

1986; Salinger, 1984; Schmalensee, 1989). One would accordingly expect a negative

relationship between the growth of an industry, and the probability of exit. The variable

INDGRSAL is derived by estimating the growth in percent of industry sales between 1981 and 1985, as reported in Trinet.

INDCONC. Industry concentration has in standard IO theory been argued to have a positive relationship with industry profitability (Bain, 1956; Porter, 1980). The reason being that scale economies and other sources of market power reduces the threat from potential entrants, allowing incumbents more room to raise prices without inviting entry. Such a relationship has found support in empirical studies (Bain, 1951; Montgomery, 1985; Weiss, 1974), but the relationship between concentration and industry profitability remains controversial, since a number of studies have failed to find such a relationship (see Schmalensee, 1989 for a review). We expect a negative relationship between the industry concentration and the probability of an industry being exited. The variable INDCONC is derived by estimating the four-firm concentration ratio of each industry for 1981, based on Trinet Data.

INDPROF. Industry profitability may be affected by numerous other factors beyond growth and concentration. To control for such unspecified factors we calculated a measure of the median return on assets for each industry over the period 1980 − 1982. The procedure used here calls for some elaboration. The Compustat database which was used to derive INDPROF consists of a segment database which report ROA in four-digit SIC codes, and a corporate database which report ROA on the firm level. We used all observations in the segment database and all single business firms in the corporate database to calculate INDPROF. However, this creates a problem because the ROA measures are not directly comparable over these two databases. Because of incomplete asset allocation, ROA is systematically higher in the segment database. To preserve observations we calculated the mean of all observations in each database for each year, and we subsequently divided each individual observation by this

mean. Thus, the individual observations were standardized as deviations in percent from the database mean for the relevant year. This allowed us to use observations from both databases and all three years, and we subsequently calculated the median of this measure for each industry. Where a minimum of five observations were obtained, this was done on the level of four digit SIC-industries, if less than five observations were obtained the same measure was calculated on the three digit level, if still less than five observations where obtained, the measure was calculated on the two digit level (following Berger and Ofek, 1995). We expect a negative relationship between INDPROF and the probability of exit.

Firm Level Independent Variables

In addition to industry level factors, we control for effects on the exit decision that are attributable to other properties of the firm than that of relatedness. We control for four firm level properties that both theory and empirical research have found to affect exit decisions. These variables are market share, parent size, parent leverage and parent liquidity.

MKTSH.   A positive relationship between market share and profitability is documented in a large number of empirical studies (e.g. Gale, 1971; Sheperd, 1972; Robins and Wiersema, 1995). There are numerous explanations for this relationship, ranging from market power explanations through cost advantages due to learning curve effects and economies of scale. Given these positive performance effects we expect firms to be less likely to exit a business where they hold large market shares. The variable MKTSH is measured as firm sales in industry *i* as percent of industry sales in 1981, and is expected to be signed negatively. The data are based on the Trinet files.

PARSIZE. The size of the parent has as noted been used as an indicator of market power and economies of scale. In addition parent size is an indicator of a parent's level of financial- and other resources. Based on this we expect a negative relationship between parent size and the probability of exit. The variable PARSIZE is measured as the total sales of the parent in 1981, based on Trinet data.

PARLEV. A highly leveraged parent may be under pressure from banks and investors to sustain a high cash flow in the short term to avoid an excessive bankruptcy risk. This is likely to reduce the patience with low performing businesses in the portfolio. Furthermore, a highly leveraged firm may experience constraints in funding attractive investment opportunities. Divesture of one business may therefore be an attractive way to finance investments in another. Therefore we expect the leverage of the parent to be positively related to the probability of a business being exited. The variable PARLEV is measured as long term debt to market value in 1981. The data were obtained from the Compustat database.

PARLIQ. A low current ratio may -like high leverage- indicate financial constraints that makes exit more likely in order to reduce bankruptcy risk, or undertake divesture as a method of financing investments in other businesses. We therefore expect the current ratio to be negatively related to the probability of a business being exited. The variable PARLIQ is measured as the ratio of current assets to current liabilities in 1981. Data were obtained from the Compustat database.

Relatedness Variables

The survivor-based relatedness measures employed in the test of H1-H3 have been presented above. The measures used in testing H1 is called $SURVTOT_i$ and $SICTOT_i$ and capture the sales weighted average relatedness of the business i to *all other* businesses in the parent k. We expect $SURVTOT_i$ to perform better than $SICTOT_i$. Note that the prefix "SURV" indicates a survivor-based measure, while the prefix "SIC" indicates a SIC-based measure. The measures used in testing H2 is called $SURVCORE_i$ and $SICCORE_i$ and capture the sales wheighted relatedness between the business i and the core business c of the parent. We expect $SURVCORE_i$ to perform better than $SICCORE_i$. In testing H3 we include all four of these measures, and we hypothesize that the two measures with the prefix "SURV" will perform better than both measures with the prefix "SIC".

Dependent Variable

The dependent variable used to test H1 and H2 is dichotomous. If a parent active in a four-digit SIC code in 1981 has exited this business by 1985, the dependent variable is given a value of 1. If the parent is still active in the industry by 1985, the value assigned is 0. Both divestures and closures are thus considered to represent exit. The Trinet database was used to identify exits and nonexits.

Variable definitions, datasources and predicted signs are summarized in Table 1 below, while Table 2 shows the means, standard deviations and correlation coefficients for all independent variables.

Insert table 1 and 2 about here

----

6. Results

The results from the logistic regression analyses are presented in table 3. Table 3 contains five

different logistic regression models. Model 1 contains control variables only. Model 2

contains control variables plus the survivor-based measure $SURVTOT_i$. Model 3 contains

control variables plus the equivalent SIC-based measure $SICTOT_i$. Model 4 contains control

variables plus the survivor-based measure $SURVCORE_i$. Model 5 contains control variables

plus the equivalent SIC-based measure $SICCORE_i$.

Insert table 3 about here

As shown in table 3 all four relatedness measures are negatively signed and significant, but

we note that while the two SIC-based measures are significant at the 0.05 level, both survivor-

based measures are significant at the 0.001 level. Looking at the Wald statistics we also note

that the survivor-based variable $SURVTOT_i$ has the highest level of significance of all the

variables included in this study.

H1 predicted that model 2 would explain the probability of exit significantly better than

model 3. This hypothesis is strongly supported. All measures of model performance improve

substantially when the measure $SURVTOT_i$ is substituted for $SICTOT_i$. The model chi-

square improves by 62.88, an improvement which is significant at the 0.001 level. The two

pseudo r^2 measures increase by 48% (Cox and Snell r^2) and 47.3% (Nagelkerke r^2) respectively. We interpret this as support for H1.

H2 predicted that model 4 would explain the probability of exit significantly better than model 5. This hypothesis is also strongly supported. All measures of model performance improve substantially when the measure $SURVCORE_i$ is substituted for $SICCORE_i$. The model chi-square improves by 33.78, again the improvement is significant at the 0.001 level. The two pseudo r^2 measures increase by 21.7% (Cox and Snell r^2) and 21.6% (Nagelkerke r^2) respectively. We interpret this as support for H2.

H3 predicted that both survivor-based measures would perform significantly better than both SIC-based measures. This hypothesis is supported if the lowest performing survivor-based measure performs significantly better than the best performing SIC-based measure. As shown in table 3 the lowest performing survivorbased measure is $SURVCORE_i$ while the best performing SIC-based measure is $SICCORE_i$. The test of H3 therefore reduces to the same test as H2, which was strongly supported. We therefore conclude that H3 is also strongly supported.

Another point worthy of mentioning is that the measure $SURVTOT_i$ performs better than the measure $SURVCORE_i$. The increase in model chi-square when moving from $SURVCORE_i$ to $SURVTOT_i$ is 6.97 which is significant at the 0.01 level. In our opinion this indicates that there are positive effects of relatedness that extends beyond the relationship between a business and the core business of the parent, and that the noted differences between model 2 and 4 reflect this. More specifically, $SURVCORE_i$ performs below $SURVTOT_i$ because WARN restricts relatedness effects to the core business in the parent portfolio.

While H1 - H3 did receive strong support, some of the control variables did not behave as expected. This was particularly the case for industry concentration (INDCONC), which was expected to be negatively related to the probability of exit, but instead a significant (though

marginally) positive relationship was found. Although surprising, this finding does have precedents (Schmalensee, 1989). One reason may be that concentrated industries are difficult to enter, and that some of the exits in our sample are in fact unsuccessful entry attempts. Another possibility is that concentrated industries are less glamorous than they are rumored to be. The conditions that cause concentration may also cause intense rivalry. For example the presence of substantial fixed costs will typically induce concentration, but it also tends to make battles for market share more intense. Thus episodes of intense competition may break out which hurt all incumbents, but smaller firms in particular. As a result these smaller firms may exit by closing down, by divesture to a larger firm, or by merger with other small firms to gain economies of scale. To test this we split the sample in two based on market share, which resulted in a significant positive relationship between concentration and probability of exit for the sub sample with the smallest market shares, and no relationship for the subsample with the largest market shares.

Another surprising finding was the insignificant relationship between industry profitability and the probability of exit. There may be several factors contributing to this finding. One reason may associated with what Goold et al. (1994) called the parenting advantage criterion. This criterion states that a business should be sold when the current owners are not the best possible owners, because the better owners will be willing to pay more than the current ones can expect from continued ownership. Abiding by this criterion would weaken the tendency to retain a business because it is profitable or in a profitable industry. Secondly, the distribution of profitability in high return industries may be highly skewed, such that even if the profitability of the median firm is high, it covers a tale of low profitability firms. Unfortunately the Compustat data used to construct the industry profitability measures in this study are to coarse to explore this possibility further. Thirdly, our results may simply be an artifact of these limitations in the underlying data.

Finally we may note that parent size, parent leverage and parent liquidity did not reach significance either. This is somewhat less surprising since the number of parents included in the final sample was 70. Thus we may lack sufficient statistical power to capture effects of the sizes associated with these variables.

7. Conclusions and Caveats

In sum, our findings are that the survivor-based measures outperform equivalent SIC-based measures in terms of explaining the probability of a business being exited. Our findings also indicate support for the relatedness hypothesis, since our data shows that relatedness does reduce the probability of a business being exited. However, in drawing these conclusions some caution is warranted.

One reason for caution is that exit may be influenced by other factors then efficiency. Some would argue that the early 1980s was a period where refocusing and de-conglomeration was increasingly fashionable, and that the exits decisions observed here may be more influenced by fashion and herd behavior than efficiency. This is both a threat to interpreting our findings as support for the relatedness hypothesis, but more importantly, it can be the case that the survivor-based measures perform better than the SIC-based measures because they capture such herd behavior better. Our data can not rule out this possibility, but a weak indication against a pure herd behavior interpretation is the fairly high positive correlation between the relatedness measures and market share (ranging from 0,20 to 0,37).

Another alternative interpretation is that the survivor-based measures outperform the SIC based measures because of contamination from motives associated with multipoint competition and mutual forbearance. This refers to a tendency among firms competing across several markets to instigate a balance of terror where competition is less aggressive than it

would have been otherwise (Edwards, 1955; Gimeno, 1999; Greve and Baum, 2001). Firms may refrain from exiting a weak position in one industry because maintaining this position is important to protect gains from mutual forbearance in another. Or in other words: the firm remains in the weak business not because it is reaping gains from relatedness, but because remaining there is a perquisite for low levels of rivalry elsewhere. The way the survivor-based measures are constructed makes them particularly well suited to capture such motives, since they are indeed built from a count of frequencies of multi market contact. SIC-based measures are not equally sensitive to these motives, because they are not constructed on the basis of multi market contact. One may therefore speculate that the superior ability of the survivor-based approach to explain exit is not due to relatedness, but to a particular sensitivity to the motives of creating and exploiting benefits from mutual forbearance.

Future work should focus on examining these two alternative interpretations of our findings, as well as testing the survivor-based measures against other dependent variables which relatedness is supposed to affect. Examples of the latter may include both patterns of entry and other performance measures than exit (e.g. ROA, Tobins q, sales growth, etc.). And of course it would be desirable with replications using data from other periods and other places. These important caveats not withstanding, it is our interpretation that given the current mismatch between available methodological tools and theoretical developments, the findings reported here strongly suggests that survivor-based measures can make a valuable contribution to research on how relatedness affects other variables of interest (i.e. what relatedness does).

REFERENCES

Alchian AA. 1950. Uncertainty, Evolution, and Economic Theory. *Journal of Political Economy* **LVIII**: 211-221.

Anderson E. 1985. The Salesperson as Outside Agent or Employee: A Transaction Cost Analysis. *Marketing Science* **4**(3): 234-254.

Bain JS. 1951. Relation of Profit Rate to Industry Concentration: American Manufacturing 1936-1940. *The Quarterly Journal of Economics* **65**(3): 293-324.

Berger PG, Ofek E. 1995. Diversifications Effect on Firm Value. *Journal of Financial Economics* **37**(1): 39-65

Caves RE, Porter ME, Spence AM. 1980. *Competition in the open economy : a model applied to Canada*. Harvard University Press: Cambridge, Mass.

Chatterjee S, Wernerfelt B. 1991. The Link between Resources and Type of Diversification - Theory and Evidence. *Strategic Management Journal* **12**(1): 33-48.

Christensen JF, Foss NJ. 1997. Dynamic corporate coherence and competence based competition: Theoretical foundations and strategic implications. In A Heene, R Sanchez (Eds.), *Competence-based Strategic Management*. Wiley: Chichester.

Eisenhardt KM. 1985. Control - Organizational and Economic Approaches. *Management Science* **31**(2): 134-149.

Elster J. 1989. *Nuts and Bolts for the Social Sciences*. Cambridge University Press: New York.

Fan JPH, Lang LHP. 2000. The measurement of relatedness: An application to corporate diversification. *Journal of Business* **73**(4): 629-660.

Farjoun M. 1994. Beyond Industry Boundaries - Human Expertise, Diversification and Resource-Related Industry Groups. *Organization Science* **5**(2): 185-199.

Foss NJ, Christensen JF. 2001. A market-process approach to corporate coherence. *Managerial and Decision Economics* **22**: 213-226.

Friedman M. 1953. The Methodology of Positive Economics. *Essays in Positive Economics*: 328. University of Chicago Press: Chicago.

Gale BT. 1971. Market Share and Rate of Return. *Review of Economics and Statistics* **54**: 412-423.

Goold M, Campbell A, Alexander M. 1994. *Corporate-level strategy : creating value in the multibusiness company*. J. Wiley: New York.

Hodgson G. 1993. *Economics and Evolution*. University of Chicago Press: Ann Arbor.

Hoskisson RE, Hitt MA. 1990. Antecedents and Performance Outcomes of Diversification - a Review and Critique of Theoretical Perspectives. *Journal of Management* **16**(2): 461-509.

Jaquemin A, Berry C. 1979. Entropy Measure of Diversification and Corporate Growth. *Journal of Industrial Economics* **27**: 359-379.

Joskow PL. 1985. Vertical Integration and Long Term Contracts: The Case of Coal-burning Electric Generating Plants. *Journal of Law, Economics and Organization* **1**(1): 33-80.

Kwoka JE, Ravenscraft DJ. 1986. Cooperation V Rivalry - Price-Cost Margins by Line of Business. *Economica* **53**(211): 351-363.

Laursen K, Meliciani V. 2000. The importance of technology-based intersectoral linkages for market share dynamics. *Weltwirtschaftliches Archiv-Review of World Economics* **136**(4): 702-723.

Levin RC, Klevorick AK, Nelson RR, Winter SG. 1987. Appropriating the Returns from Industrial-Research and Development. *Brookings Papers on Economic Activity*(3): 783-831.

Markides CC, Williamson PJ. 1994. Related Diversification, Core Competences and Corporate Performance. *Strategic Management Journal* **15**: 149-165.

Markides CC, Williamson PJ. 1996. Corporate diversification and organizational structure: A resource-based view. *Academy of Management Journal* **39**(2): 340-367.

McFadden D, Manski CF. 1981. Alternative estimators and sample designs for discrete choice analysis. In D McFadden, CF Manski (Eds.), *Structural analysis of discrete data with econometric applications*: xxv, 477 s. MIT Press: Cambridge, Mass.

Milgrom P, Roberts J. 1992. *Economics, organization and management*. Prentice-Hall: Englewood Cliffs, N.J.

Monteverde K, Teece DJ. 1982. Supplier Switching Costs and Vertical Integration in the Automobile-Industry. *Bell Journal of Economics* **13**(1): 206-213.

Montgomery CA. 1985. Product-Market Diversification and Market Power. *Academy of Management Journal* **28**(4): 789-798.

Montgomery CA, Hariharan S. 1991. Diversified Expansion by Large Established Firms. *Journal of Economic Behavior & Organization* **15**(1): 71-89.

Penrose E. 1959. *The theory of the growth of the firm*. Blackwell: Oxford.

Piscitello L. 2000. Relatedness and coherence in technological and product diversification of the world's largest firms. *Structural Change and Economic Dynamics* **11**: 295-315.

Reed R, Luffman GA. 1986. Diversification: The Growing Confusion. *Strategic Management Journal* **7**: 29-36.

Richardson GB. 1972. The Organisation of Industry. *Economic Journal* **82**(327): 883-896.

Robins J, Wiersema MF. 1995. A Resource-Based Approach to the Multibusiness Firm - Empirical-Analysis of Portfolio Interrelationships and Corporate Financial Performance. *Strategic Management Journal* **16**(4): 277-299.

Robins JA, Wiersema MF. 2003. The measurement of corporate portfolio strategy: Analysis of the content validity of related diversification indexes. *Strategic Management Journal* **24**(1): 39-59.

Rumelt RP. 1974. *Strategy, structure, and economic performance*. Division of Research Graduate School of Business Administration Harvard University ; distributed by Harvard University Press: Cambridge, Mass.

Rumelt RP. 1982. Diversification Strategy and Profitability. *Strategic Management Journal* **3**(4): 359-369.

Salinger MA. 1984. Tobin-Q, Unionization, and the Concentration-Profits Relationship. *Rand Journal of Economics* **15**(2): 159-170.

Schmalensee R. 1989. Inter-Industry Studies of Structure and Performance. In R Schmalensee, RD Willig (Eds.), *Handbook of Industrial Organization*. North Holland: New York.

Sharma A. 1998. Mode of Entry and Ex - Post Performance. *Strategic Management Journal* **19**: 879-900.

Sheperd WG. 1972. The Elements of Market Structure. *Review of Economics and Statistics* **54**: 25-37.

Silverman BS. 1999. Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. *Management Science* **45**(8): 1109-1124.

Teece DJ. 1980. Economies of Scope and the Scope of the Enterprise. *Journal of Economic Behavior & Organization* **1**(3): 223-247.

Teece DJ. 1982. Towards an Economic-Theory of the Multiproduct Firm. *Journal of Economic Behavior & Organization* **3**(1): 39-63.

Teece DA. 1986. Profiting from technological innovation. *Research policy* **15**: 286-305.

Teece DJ, Rumelt R, Dosi G, Winter S. 1994. Understanding Corporate Coherence - Theory and Evidence. *Journal of Economic Behavior & Organization* **23**(1): 1-30.

Voight CW. 1993. The Trinet large establishment data. University of california at Los Angeles, CA: Anderson Graduate School of Management.

Weiss LV. 1974. The concentration-profits relationship and antitrust. In Goldschmidt et al. (Ed.), *Industrial Concentration: The New Learning*: 201-220. Little, Brown: Boston, MA.

Williamson OE. 1985. *The economic institutions of capitalism : firms, markets, relational contracting.* Free Press: New York.

Willig RD. 1979. Multiproduct Technology and Market Structure. *American Economic Review* **69**(2): 346-351.

Winter SG. 1971. Satisficing, Selection, and Innovating Remnant. *Quarterly Journal of Economics* **85**(2): 237-261.

Table 1: Variable definitions, datasources and predicted signs

Variable	Definition	Data Source	Predicted Sign
INDGRSAL	Salesgrowth in industry i between 1981 and 1985	Trinet	-
INDCONC	4-firm concentration ratio in 1981 in industry i	Trinet	-
INDPROF	Industry median ROA 1980-1982 in industry i	Compustat	-
MKTSH	Market share in industry i for the parent in 1981	Trinet	-
PARSIZE	Total sales of the parent in 1981	Trinet	-
PARLEV	Leverage of the parent in 1981	Compustat	+
PARLIQ	Current ratio of the parent in 1981	Compustat	-
SURVTOT	Weighted average *survivor-based* relatedness of industry i to all other industries in the portfolio of the parent	Trinet	-
SICTOT	Weighted average *SIC-* based relatedness of industry i to all other industries in the portfolio of the parent	Trinet	-
SURVCORE	Salesweighted *survivor-based* relatedness of industry i to the core business of the parent	Trinet	-
SICCORE	Salesweighted *SIC*-based relatedness of industry i to the core business of the parent	Trinet	-

Table 2   Means, Standard Deviations and Correlation Coefficients of Independent Variables   N=1191

	Variables	Mean	SD	Indgrsal	Indconc	Indprof	Mktsh	Parsize	Parlev	Parliq	Survtot	Sictot	Survcore
	Industry variables												
1	Indgrsal	0,335	0,605										
2	Indconc	27,24	18,08	0,16***									
3	Indprof	0,064	0,384	0,16***	-0,03								
	Firm variables												
4	Mktsh	2,22	5,49	0,14***	0,45***	-0,09***							
5	Parsize	63713	62907	-0,02	0,10***	-0,01	0,14***						
6	Parlev	0,179	0,094	0,03	-0,04	-0,07**	-0,08**	-0,25***					
7	Parliq	1,64	0,49	-0,05	-0,08***	0,03	-0,03	-0,42***	-0,08***				
	Relatedness variables												
8	Survtot	8,69	6,77	0,09***	0,09***	0,04	0,20***	-0,03	0,06**	0,00			
9	Sictot	0,129	0,244	0,12***	0,13***	0,09***	0,25***	-0,06**	0,08***	0,06**	0,53***		
10	Survcore	1,21	6,42	0,03	0,09***	-0,04	0,29***	-0,02	0,04	-0,06**	0,38***	0,25***	
11	Siccore	0,04	0,23	0,05*	0,13***	-0,04*	0,37***	-0,00	0,04	-0,05**	0,31***	0,29***	0,88***

*** Correlation is significant at the 0.01 level

** Correlation is significant at the 0.05 level

* Correlation is significant at the 0.1 level

Table 3: Logistic Regression Output   N = 1191

	Model 1 Ctrls Only		Model 2 Ctrls + SURVTOT		Model 3 Ctrls + SICTOT		Model 4 Ctrls + SURVCORE		Model 5 Ctrls + SICCORE		Signs	
	Beta	Wald	Beta	Wald	Beta	Wald	Beta	Wald	Beta	Wald	Expected	Observed
Constant	0,44	1,60	1,02***	7,69	0,45	1,64	0,77**	4,56	0,57*	2,62		
Indgrsal	-0,69***	30,61	-0,68***	28,09	-0,68***	29,19	-0,70***	30,70	-0,70***	31,01	-	-
Indconc	0,01**	3,18	0,01*	3,13	0,01**	3,22	0,04	1,07	0,01*	2,22	-	+
Indprof	-0,06	0,15	-0,02	0,01	-0,04	0,06	-0,14	0,73	-0,13	0,59	-	
Mktsh	-0,15***	37,39	-0,12***	25,51	-0,14***	32,89	-0,08***	10,56	-0,11***	21,34	-	-
Parsize	0,00	0,50	0,00	1,12	0,00	0,67	0,00*	2,45	0,00	1,30	-	
Parlev	-0,61	0,78	-0,33	0,21	-0,51	0,54	-0,53	0,56	-0,46	0,44	+	
Parliq	0,01	0,00	0,04	0,09	0,03	0,03	-0,06	0,18	-0,05	0,11	-	
SURVTOT$_i$			-0,09***	53,63							-	-
SICTOT$_i$					-0,53**	3,38					-	-
SURVCORE$_i$							-1,18***	16,59			-	-
SICCORE$_i$									-6,45**	3,69	-	-
-2Log likelihood	1531,46		1465,10		1527,98		1472,06		1505,83			
ModelChi-square	119,59***		185,96***		123,08***		178,99***		145,21***			
Δ Chi-square vs. model 1			66,37***		3,49**		59,40***		25,62***			
Δ Chi-square model 2 vs. model 3					62,88***							
Δ Chi-square model 4 vs. model 5									33,778***			
Cox and Snell r^2	0,096		0,145		0,098		0,140		0,115			
Nagelkerke r^2	0,127		0,193		0,131		0,186		0,153			

*** Correlation is significant at the 0.01 level
** Correlation is significant at the 0.05 level
* Correlation is significant at the 0.1 level

Relatedness and Patterns of Diversification:
A Survivor-based Approach

Lasse B. Lien

Institutt for strategi og ledelse

Norges Handelshøyskole

Breiviksveien 40

N-5045 Bergen

Norway

lasse.lien@nhh.no

Paper no. 3

Relatedness and Patterns of Diversification: A Survivor-based Approach

Abstract

This study addresses the impact of relatedness on patterns of corporate diversification. More specifically we compare a survivor-based approach to measuring relatedness with the conventional SIC-based approach in terms of explaining which industries a firm diversifies into. Our findings show that the survivor-based approach is substantially superior in terms of explaining entry decisions by diversified firms. These findings complement the recent observations that survivor-based measures of relatedness outperform SIC-based measures in terms of explaining exit decisions by diversified firms, and thus strengthen the claim that survivor-based measures of relatedness can make a valuable contribution to research on diversification and corporate strategy.

1. Introduction

No variable within research on diversification and corporate strategy has been given an amount of attention greater than that of relatedness. A sizable share of this attention concerns how relatedness should be measured (Caves, Porter and Spence, 1980; Chatterjee and Blocher, 1992; Fan and Lang, 2000; Farjoun, 1994; Hoskisson et al., 1993; Jacquemin and Berry, 1979; Markides and Williamson, 1994, 1996; Montgomery and Hariharan, 1991; Robins and Wiersma, 1995, 2003). Several authors have pointed to a gap between theoretical conditions and the way relatedness is operationalized, and suggest this as the main cause of inconsistencies between theoretical predictions and empirical findings in the existing literature (Markides and Williamson, 1994, 1996; Robins and Wiersma, 1995). In practice, however, this gap is extremely difficult to close.

Recently, Lien (2003) suggested a different approach towards narrowing this gap. Building on work by Teece et al. (1994), he tested a survivor-based approach to measuring relatedness, and found that survivor-based measures performed significantly better than SIC-based measures in terms of explaining the *exit* decisions of diversified firms (exit was used as a performance measure). This study extends this work by comparing survivor-based measures with SIC-based measures in terms of explaining *entry* decisions.

A survivor-based approach to measuring relatedness means that what is related is what firms in competitive markets combine often. More specifically it implies that a measure of the relatedness between a pair of industries can be obtained by considering how often they are actually combined inside a firm - compared to what one would expect if diversification patterns were random. Industries are related when this difference is large and positive, and they are unrelated if it is negative. This line of thinking was originally suggested by Teece, Rumelt, Dosi and Winter (1994), but these authors only used it to illustrate that coherence

101

(non-randomness) was a salient attribute of the diversification patterns of US firms. Apart from the noted study by Lien (2003) - which focused on exit decisions - the procedure has to our knowledge not been evaluated as a method of capturing inter-industry relatedness.

We proceed as follows: Section 2 discusses whether entry decisions can be used to compare measures of relatedness. Section 3 discusses the limitations of the most frequently used relatedness measures. Section 4 presents an alternative procedure based on the survivor logic. Two variants of a survivor-based measure of relatedness are then developed, and hypotheses contrasting these with equivalent measures based on the SIC-system are formulated. Section 5 discusses methodological issues. Section 6 presents our empirical findings, and section 7 concludes.

2. Can Entry Decisions be Used to Compare Measures of Relatedness?

This study relies in a crucial way on three postulates:

1) Relatedness matter for performance
2) Performance matter for decision makers
3) Decision makers have relevant information about relatedness

Based on these postulates we suggest that actions taken by decision makers (i.e. entry decisions) contain information about true relatedness, and accordingly, that relatedness measures can be compared by their relative ability to explain actual entry decisions. Our motive for formulating these postulates is that entry decisions have not been screened by the competitive process. Therefore we cannot rely on the competitive process to sort between efficient and inefficient decisions. In contrast, if our dependent variable was a performance measure, such as exit or ROA, or any other variable that is measured after the competitive process has been set to work, we would not necessarily have to resort to postulate 2 and 3. If

our belief in the effectiveness of the competitive process was sufficiently strong, we could

rely on this process to produce outcomes *as if* these postulates were empirically true (Alchian,

1950; Friedman, 1953). However, since our dependent variable is *ex ante* to competition, we

cannot make such a claim.[31] Yet, if data on actual entry decisions are to be of use as a way of

comparing relatedness measures, all three postulates must be included. We shall therefore

devote some space to arguing why we consider each of the three postulates plausible (one

should also note that there are several widely cited precedents for using patterns of entry as

evidence of the beneficial effects of relatedness, e.g. Farjoun, 1994; Montgomery and

Hariharan, 1991; Silverman, 1999).

2.1 Relatedness and Performance

The classic "economies of scope" view of why and when relatedness matters for performance

involves three conditions. The first is that the entrant possesses some resources that are

functional substitutes for resources in the target industry (below we expand the discussion to

include situations involving complementarity). This condition ensures that the entrant

possesses some resources that are relevant in the destination industry.

The second is that there are some indivisibilities associated with these resources, so that the

entrant possesses some excess capacity. This criterion is usually attributed to Penrose (1959).

Penrose highlighted that firms in the course of their normal operations, as a result of learning,

continuously generated new resources - and excess capacity in existing resources. To the

extent that resources are not fully exploited in existing businesses - they can be deployed at

low marginal cost in a new business. Conversely, if no excess capacity exists, a diversified

firm would have no efficiency advantages over an independent start up, and probably some

---

[31] Note, however, that the key independent variables (the survivor-based relatedness measures) are ex post to
screening by the competitive process.

disadvantages vis a vis incumbents in the destination industry (assuming some entry barriers exist).

The third is that there must exist some form of market failure for trading in this excess capacity (Teece, 1980, 1982). If the relevant excess capacity could be sold or rented out in well functioning markets, this would probably be preferable to full scale diversification. Full scale diversification will normally involve some degree of weakened incentives, increased complexity, added overhead, and a need to acquire additional resources at full cost. However, the presence of some form of market failure can neutralize these concerns because missing markets and high transaction costs may make market contracting unfeasible as a way to capitalize on excess capacity (Teece, 1980, 1982; Williamson, 1985).

More recently, several authors have argued for a shift in focus from resource substitutability and excess capacity, to resource complementarity (Christensen and Foss, 1997; Foss and Christensen, 2001; Teece et al., 1994). Complementarity refers to situations where there are positive externalities (which essentially is a different form of indivisibility) across industry boundaries. For example because the value of assets in one industry is affected positively by the level and uses of assets in another. This creates a quantitative and qualitative coordination problem, and a firm may diversify to better exploit such positive externalities (Richardson, 1972).[32] However, this requires that a diversified firm can solve this coordination problem in ways that a single business firm cannot, which in turn requires that some form of market failure exists (Milgrom and Roberts, 1992). Hence, the condition of market failure also applies in situations involving complementarity.

In sum, any kind of similarity between industries cannot be expected to bring about a positive relationship between relatedness and performance. Indivisibilities (either in the form of

---

[32] Or to explore the existence of such complementarities and develop new ones (Foss and Christensen, 2001; Klein and Klein, 2001).

productive capacity or positive externalities) must exist, and market contracting must not leave a single business firm or an unrelated diversifier equally well off. Unfortunately, as will be discussed in section 3, developing measures of relatedness that screens for these conditions in a convincing manner is extremely difficult.

2.2 Performance and Decision Makers

We believe that those making diversification decisions are very much concerned with the performance outcomes of their diversification decisions. Diversification decisions are highly visible decisions, which if proven unsuccessful could damage both the compensation, job security and reputation of those responsible. We certainly acknowledge that decision makers may have other goals than economic performance (Jensen and Meckling, 1976), including maximizing the size of the firm (Mueller, 1969), excessive reductions in bankruptcy risk (Amihud and Lev, 1981), and more. However, two reasons lead us to believe that expected performance remains the dominant criterion. First of all a manager maximizing growth or some other managerial benefit is likely to choose the most profitable opportunities first (Montgomery, 1994). Secondly, the early nineteen eighties, which is the time frame focused in this study, was a period where unrelated diversification was becoming increasingly "unfashionable" (Shleifer and Vishny, 1994). Thus, it seems reasonable that the pursuit of unrelated diversification was both more likely to be met with resistance from board members and owners, and more likely to be costly in terms of negative responses from capital markets and other stakeholders, than earlier periods.

In sum, while acknowledging that some diversification moves may be influenced by other motives than expected performance, we believe that expected performance is the dominant criterion used by decision makers. Thus, the decisions they make contain information about what they believe is the best course of action in terms of expected performance. However, we

stress that since our task is to compare different measures of relatedness, it is much more important that the different relatedness measures we compare do not correlate systematically different with such other motives, than it is to conclude that such motives are insignificant.

2.3   Decision Makers Have Relevant Information About Relatedness

Decision makers making entry decisions are human beings, and therefore subject to bounded rationality and various biases. Nevertheless, we believe that these decision makers do not act randomly, and that such high profile and high commitment decisions as the entry into a new industry are usually the subject of considerable analysis and reflection. Therefore we believe that the entry decisions they make contain information - not only about perceived relatedness - but also about "true" relatedness. Again, however, what is important is not that biases and bad judgements are absent, but that these do not correlate systematically different with the relatedness measures we compare.

3. Problems with the Existing Measures of Relatedness

The most frequently used measures of relatedness in existing research are categorical measures and continuous SIC-based measures. As argued below there are problems associated with both of these measures, in particular with their ability to capture all the conditions specified in section 2.1.

Categorical Measures

The categorical approach is dominated by the work of Rumelt (1974) which has become the standard for the use of categorical measures. Based on three ratios, Rumelt classified diversification strategies into four broad categories. These were: single business firms,

dominant business firms, related firms, and unrelated firms. The ratios used for classification were:

- *Specialization ratio*: The proportion of a firm's revenue that can be attributed to its largest single business
- *Related ratio*: The proportion of a firm's revenue that can be attributed to its largest group of related businesses
- *Vertical ratio*: The proportion of a firm's revenue that arise from all byproducts, intermediate products, and end products of a vertically integrated sequence of processing activities

An important element is the definition of what constitutes a related business, which subsequently affects the related ratio. This was done subjectively using similarities in inputs, production technology, distribution channels and customers. As many others have pointed out, there are potential problems with this procedure. One is of course the reliability- and intersubjectivity problems associated with the subjective element in the classifications, another is that by measuring relatedness on a nominal level it only allows for comparisons of within group averages and the procedure therefore becomes quite restrictive.

These concerns not withstanding, an equally important problem is that the procedure mainly seems to capture the degree to which one can assume that resources are potential substitutes across industry boundaries. This will probably be the case when there are "large similarities in inputs, production technology, distribution channels and customers". The procedure looms silent on the likelihood that such resources are imperfectly divisible, and hence whether excess capacity is likely to develop, nor does it contain any notion of market failure that make contracting over the relevant resources costly. Both of these conditions were above identified as necessary conditions for economies of scope to create an advantage for a related diversifier, hence such a measure would be prone to exaggerate relatedness in some instances (i.e. where resources are close substitutes, but these additional conditions are not met). The implicit focus on similarities and economies of scope also raises the concern that such a procedure may not capture complementarity well, which implies that it will underestimate relatedness in other

instances (Foss and Christensen, 2001). Given these shortcomings vis a vis the theoretical

conditions, categorical measures have important limitations in terms of isolating the

theoretically salient types of relatedness.

Continous SIC-based Measures

The continous SIC-based measures are currently the most widely used approach (Robins and

Wiersma, 2003), they include measures such as the entropy index (Jaquemin and Berry,

1979), and the concentric index (Caves, Porter and Spence, 1980).[33] The advantages of using

measures based on the SIC-system are that there is no subjective element in classifying the

degree of relatedness, and it also allows relatedness to be measured on an interval level. The

2, 3 and 4-digit levels in the SIC-system are treated as points on an underlying scale of

relatedness, and arithmetic values are assigned to the distances. This allows for a wide range

of statistical operations, and use of the large amounts of secondary data available in the SIC-

format.

However the use of distances in the SIC-system also introduces problems, in that it imposes

some very strong assumptions on the SIC-system. It assumes that industries are homogenous

within category levels, which is problematic if the breadth of the industry classifications vary.

---

[33] The concentric index is calculated as follows :

$$FDIVERS_k = \Sigma_i P_{ki} \, \Sigma_j P_{kj} d_{ij}$$

Where:

P_{ki} = percentage of sales for firm k in industry i

P_{kj} = percentage of sales for firm k in industry j

d_{ij} = weighting factor such that $d_{ij} = 0$ where i and j belong to the same 3-digit SIC-category, $d_{ij} = 1$ where i and j belong to the same 2-digit category but different 3-digit categories, and $d_{ij} = 2$ where i and j are in different 2-digit categories.

The related portion of entropy is calculated as follows:

$$E_R = E_T - E_U = \Sigma P_T \, ln(1/P_T) - \Sigma P_U \, ln(1/PU)$$

Where:

E_R = Related component of entropy

E_T = Entropy defined at the 4-digit level

E_U = Entropy defined at the 2-digit level

P_T = Percentage of sales in each 4-digit industry

P_U = Percentage of sales in each 2-digit industry

In fact most observers agree that they do (Robins and Wiersma, 1995; Rumelt, 1982). The second assumption is that it assumes that there is equal dissimilarity between real industries when they are equally distant within the SIC-hierarchy. This assumption is also problematic. Our major point, however, is that SIC-based measurements do not fare any better than the categorical measures in terms of capturing the conditions of indivisibility and market failure.[34] This means that even if we believe that there is a high correlation between distances in the SIC-system and the degree to which resources are potential substitutes across industries, the omission of these two conditions will imply that this kind of relatedness is prone to be exaggerated. And according to Foss and Christensen (2001) the SIC-based procedure has an implicit bias towards economies of scope, which indicates that it is not likely to capture complementarities well, suggesting that this type of relatedness is prone to be underestimated. The combined effect of over- and underestimation of relatedness is that the continuous SIC-based measures are quite crude in terms of capturing the types of relatedness theory suggest is important.

In sum, there are substantial problems with both the categorical and continuous relatedness measures. The survivor-based approach we are about to suggest is also less than perfect, but our goal is to examine whether it is less so than the alternatives.

4. The Alternative: A Survivor-based Approach to Relatedness

As indicated above, our intention is to compare continuous SIC-based measures of relatedness with survivor-based measures of relatedness in terms of explaining the entry decisions made by diversified firms. We now elaborate on the reasoning behind the survivor-based approach, and how the survivor-based measures are constructed.

---

[34] A study that does consider the condition of excess capacity is Chatterjee and Wernerfelt (1991).

The fundamental survivor-based measure relies crucially on what has become known as the survivor principle (henceforth: SP). The SP has as it core idea the notion that the competitive process screens for efficiency - and does so well enough that a sample of competitive firms will be dominated by the decisions or behaviors that are efficient - at least in a comparative sense (Alchian, 1950:211).[35] Two key processes are given the burden of ensuring this. One is that firms making negative profits will, unless some corrective measure is taken, loose resources and ultimately become extinct, while firms making positive profits will acquire resources and grow. The other is that the desire to make positive profits provides a strong incentive for the less successful firms to imitate the more successful firms. While few believe that the competitive process performs this screening perfectly[36] the behavior of researchers in the field of economics, organizational economics and strategic management indicate quite an optimistic view of this process. After all theories or hypotheses about what is efficient are routinely tested by measuring what firms actually do, which indicates a belief in the basic conjecture of the SP (i.e. that competitive markets display what is efficient).

If the SP works, a measure of relatedness can be built on the idea that what is related is what firms in competitive markets combine often. The fundamental premise of this is in other words that industries that are related will be more frequently combined within a firm. More specifically, we estimate how much the frequencies of *actual* combinations of four-digit SIC industries deviate from what one would expect if diversification patterns were *random*. We take this difference to constitute a survivor-based measure of the relatedness between a pair of industries.

The potential advantage of this survivor-based measure is that it incorporates the knowledge of the best informed actors (which presumably are those making portfolio decisions), but even if their information is poor, their decisions have been screened by the competitive process.

---

[35] For the view that the competitive process creates outcomes that are optimizing, cfr. Friedman (1953).
[36] Milton Friedman is a possible exception (Friedman, 1953).

which will enforce a reversal of poor decisions. Therefore it is not implausible that a survivor-based measure is better at capturing relatedness than the existing alternatives (note here that it is critical to distinguish between the survivor-based measures of relatedness we use, which *have* been screened by the competitive process, and the entry decisions we use those relatedness measures to analyze, which *have not* been subject to the competitive process). On the other hand, the behavior of decision makers is surely not optimal, and screening function of the competitive process is surely not perfect either. A survivor-based measure will therefore include noise. Several authors criticizing the SP, have indicated that they believe the level of noise will be substantial (e.g. Elster, 1989; Hodgson, 1993; Winter, 1971). However, in terms of measuring relatedness the existing alternatives involve substantial noise too. The question is therefore one of relative noise, and the best judge on this issue is data. We now move on to describe in detail how a survivor-based measure of relatedness can be constructed. Our approach is based on a procedure originally developed by Teece, Rumelt, Dosi and Winter (1994).

Let the universe of diversified firms consist of K firms, each active in two or more of I industries. Let $C_{ik} = 1$ if firm k is active in industry i. The number of industries participated in by firm k is $m_k = \Sigma_i C_{ik}$ and the number of diversified firms present in industry i is $n_i = \Sigma_k C_{ik}$. Let J_{ij} be the number of diversified firms active in both industries i and j, such that $J_{ij} = \Sigma_k C_{ik} C_{jk}$. Thus J_{ij} is a count of how often industries i and j are actually combined within the same firm. J_{ij} will be larger if industries i and j are related, but will also increase with n_i and n_j. To remove the effect of the size of industries i and j, the number J_{ij} is compared with the number of expected combinations if diversification patterns were random.

The random diversification hypothesis can be operationalized as a hypergeometric situation where a sample of size n_i is drawn (without replacement) from a population of K firms. Those

chosen are considered active in industry i. A second independent sample of size n_j is then

drawn from the population the population of K firms. Those chosen are considered active in

industry j. The number x_{ij} of firms active in both i and j is then a hypergeometric random

variable with population K, special members n_i and sample size n_j. The distribution function

for this variable is then:

$$\Pr(X_{ij} = x) = f_{hg}(x, K, n_i, n_j) = \frac{\binom{n_i}{x}\binom{K - n_i}{n_j - x}}{\binom{K}{n_j}}$$

The mean and variance of X_{ij} are:

$$\mu_{ij} = E(X_{ij}) = \frac{n_i n_j}{K},$$

$$\sigma^2 = \mu_{ij}\left(1 - \frac{n_i}{K}\right)\left(\frac{K}{K - 1}\right).$$

A standardized measure of the relatedness between industries i and j is then constructed based

on the difference between J_{ij} and μ_{ij} in the following fashion:

$$SR_{ij} = \frac{J_{ij} - \mu_{ij}}{\sigma_{ij}}$$

The measure SR_{ij} is thus a standardized measure of how much actual combinations exceed

expected combinations under the random diversification hypothesis. With this fundamental

measure of the relatedness between a pair of businesses it is possible to compute various

relatedness measures. The two variants chosen here reflect our choice of dependent variable,

which is the probability that a given diversified firm will enter a given industry. Thus. we

wish to develop measures that cast light on the relatedness between a candidate industry, which is subsequently either entered or not entered, and the industries present in the parent portfolio.

The first measure captures the weighted average relatedness of a potential destination industry i to all other businesses in the parent portfolio. Assume a diversified firm that participates in m industries. Its business in industry j has sales of s_j and survivor-based relatedness SR_{ij} with industry i. The weighted average relatedness SURVTOT$_i$ of the target industry i to all other business in the firm is then defined as:

$$SURVTOT_i = \frac{\sum SR_{ij} s_j}{\sum s_j}$$

A parallel measure based on SIC-distances can be obtained as follows:

$$SICTOT_i = \frac{\sum d_{ij} s_j}{\sum s_j}$$

Where $d_{ij} = 2$ if i and j are in the same 3-digit SIC codes

$d_{ij} = 1$ if i and j are in different 3-digit-, but the same 2 digit SIC codes

$d_{ij} = 0$ if i and j are in different 2-digit SIC codes

Note that this approach builds on the concentric index (Caves et al., 1980), but has been modified to accommodate our choice of dependent variable (a precedent of this modified concentric index is Sharma, 1998).

An alternative approach does not consider how related a destination industry is to *all* other business in the corporate portfolio, but how related it is to the *two closest* neighboring businesses of the parent. The approach here is to rank the survivor-based measure SR_{ij} between the destination industry i, and all other industries in the parent portfolio. The two industries with the highest measure of SR_{ij} are considered the neighboring businesses. Let λ_{ij} = 1 for a business that is defined as a neighbor to business i, and λ_{ij} = 0 for those that are not. The weighted average relatedness of neighbors to business i is then defined by:

$$SURVNBOR_i = \frac{\sum SR_{ij}s_j\lambda_{ij}}{\sum s_j\lambda_{ij}}$$

Again, we also computed a parallel measure based on SIC-distances:

$$SICNBOR_i = \frac{\sum d_{ij}s_j\lambda_{ij}}{\sum s_j\lambda_{ij}}$$

Where d_{ij} = 2   if i and j are in the same 3-digit SIC codes

d_{ij} = 1   if i and j are in different 3-digit-, but the same 2-digit SIC codes

d_{ij} = 0   if i and j are in different 2-digit SIC codes

λ_{ij} = 1   if business j is defined as a neighbor to business i[37]

λ_{ij} = 0   if business j is not defined as a neighbor to business i

---

[37] Note that we resorted to the following procedure when business were equidistant from the target industry: First we identified the closest neighbor, if several where equidistant, we used the sales of the *largest* business to compute sales weights. Then we identified the second closest. If there were several second closest firms, we chose the *smallest* of these. This was done to reflect our assumption that one would expect closer cooperation with the closest neighbor than the second closest. This procedure implies that when the closeness to the two neighbors differed, we weighted our measure in favor of the closest of the two.

An example may clarify this latter measure. Assume that the two closest related industries of a parent are one in the same 3-digit SIC code (d_{ij} = 2), and one in the same 2-digit code (d_{ij} = 1). And assume further that the relative size of these businesses (s_j) implies that they are weighted with 75% and 25% respectively. The measure SICNBOR$_i$ will then be found as follows:

$$SICNBOR_i = 2 * 0.75 + 1 * 0.25 = 1.75$$

Given these four measures: SURVTOT$_i$, SICTOT$_i$, SURVNBOR$_i$, and SICNBOR$_i$, we can formulate the following hypotheses.

H1:   SURVTOT$_i$ will explain the probability of entry significantly better than SICTOT$_i$

H2:   SURVNBOR$_i$ will explain the probability of entry significantly better than SICNBOR$_i$

H3:   SURVTOT$_i$ and SURNBOR$_i$ will both explain the probability of entry significantly better than SICTOT$_i$ and SICNBOR$_i$

Note that while H1 and H2 states that survivor-based measures of relatedness will perform better than parallel measured based on SIC-data, H3 states that both survivor-based measures will perform better than both SIC-based measures. Note also our expectation that all four measures will be positively signed.

5. Methodology

This study involves two distinct empirical operations. First we had to calculate the fundamental survivor-based measure of relatedness SR_{ij} for all possible pairs of industries in the US economy. With this measure in hand we were able to calculate the survivor-based measures $SURVTOT_i$ and $SURVNBOR_i$ for any specific business belonging to any specific parent. The second empirical operation is to conduct a test of our hypotheses linking these two measures and their SIC-based equivalents to the probability of a given parent entering a given industry.

5.1 Calculating SR_{ij}

To calculate SR_{ij} we used the AGSM/Trinet Large Establishment Database (Trinet). The Trinet database contains records of all US establishments with more than 20 employees.[38] including variables such as four-digit SIC code, corporate ownership and sales. By aggregating the establishments for each parent in each four digit sic-code, and the different four digit sic-codes for each parent, and different parents for each four-digit SIC industry, we are able to get a comprehensive picture of diversification patterns in the US-economy. Comparison with the Census of Manufacturers indicate that Trinet contains 95% of all establishments it should (Voight, 1993), and that omissions are most likely for small firms (which are less likely to be diversified). The primary measure of SR_{ij} was calculated from the Trinet files of 1981, using all recorded firms active in two or more four-digit SIC-codes as a basis. After deleting single business firms, government owned and non-profit industries, this resulted in a total of 13.164 diversified firms, active in 929 different industries, covering a total of 57.647 individual businesses. Of the 431.056 possible industry pairs, 122.105 were

---

[38] Trinet also includes foreign establishments in the US

observed. The measure of SR_{ij} between these ranged from -7.97 to 93.55 with a mean of 4.33

and a standard deviation of 5.06. But note that we also calculated SR_{ij} for industry pairs that

where not combined by 1981, because some of these where combined in the subsequent

periods where we observed entry or nonentry, and because some of the randomly chosen non-

entries created a need to calculate SR_{ij} for industry pairs that were never combined. However,

it is not in any way problematic to calculate this measure for unobserved combinations. All

this implies is that the number of observed combinations (J_{ij}) is set to zero in the formula for

calculating SR_{ij}. Based on these calculations of SR_{ij}, we were able to calculate measures of

SURVTOT$_i$ and SURVNBOR$_i$ by following the procedures described in section 4.

5.2 Testing H1-H3

Sample

The sample for testing H1-H3 was derived as follows. We started out with all the 13164

diversified firms in the Trinet database. To obtain the necessary data for all the variables of

interest, Trinet data had to be merged with financial data from Compustat database. Since the

parent identity numbers in these two databases are different, the matching had to be done

alphanumerically by parent name. Spelling differences between the two databases resulted in

undisputable matches for 854 companies that had entries in all the years needed to compute

the variables. We believe this matching procedure to be a functional equivalent of random

sampling, since there is no reason to expect spelling matches of parent names in the two

databases to be biased in any particular way.[39]

We imposed three further restrictions on our sample. One was to remove all firms with sales

below $ 10 million dollars, the second was to remove firms that where sold or liquidated in

their entirety between 1981 and 1985. The third and final criterion was that the firm had

---

[39] The matching of Trinet and Compustat data creates a bias toward larger firms compared to Trinet data alone. The reason for this is that Compustat contains publicly traded firms only, while Trinet contains both. Publicly traded firms are on average larger than privately held firms.

entered at least one new four digit SIC-code between 1981 and 1985. This reduced the sample to 145 firms. These 145 firms operated 4582 businesses in 724 different four digit SIC-codes in 1981. They entered a total of 1202 industries between 1981 and 1985, exited 1118 industries, and remained in 3464 industries throughout the period.

To test our hypotheses we included all the 1202 instances of entry, but rather than using the entire sample of non-entries, we used the random sample generator in SPSS to select a sample of non-entries of comparable size. State based sampling has been suggested as preferable to a pure random sample when a population is overwhelmingly characterized by one state, and will provide unbiased and consistent coefficients for all variables except the constant term (McFadden and Manski, 1981). Adjusting for 26 cases for which data were missing, this resulted in a final sample of a total of 2378 observations, 1202 of which were entries, and 1176 of which were non-entries.

Statistical Methods

In order to test H1-H3 we developed a model of the relationship between the probability of entry and relatedness, which controls for industry- and parent variables that previous research indicates may affect entry decisions. This was done to control for other potential motives for entry than to reap the gains from relatedness, such as market power or joining a high growth or high return industry, and to reduce the risk that the performance of one or more of the presented relatedness measures are inflated or deflated because of associations with "other" variables. Given the dichotomous nature of the dependent variable, a logistic regression analysis was considered appropriate for testing our model. The general model is the following:

$P(\text{Entry}=1) = \beta_1 + \beta_2(\text{Industry growth}) + \beta_3(\text{Industry concentration}) + \beta_4(\text{Industry}$

$\text{profitability}) + \beta_5(\text{Parent size}) + \beta_6(\text{Parent diversity}) + \beta_7(\text{Parent relatedness}) + \varepsilon$

Industry Level Independent Variables

INDGRSAL. Industry growth is widely assumed to affect industry attractiveness favorably,

because it allows firms to grow without having to steal customers from competitors. Thus,

industry growth tends to soften competitive rivalry and raise the average profitability. Such a

relationship has been confirmed in numerous empirical studies (i.e. Kwoka and Ravenscraft,

1986; Salinger, 1984; Schmalensee, 1989). High growth may therefore function as a substitute

for close relatedness in the eyes of a decision maker who is concerned with post entry

performance. In addition, decision makers may as discussed earlier obtain private benefits

from growth, which may create a bias towards entering high growth industries. For these

reasons one would expect a positive relationship between the growth of an industry, and the

probability of entry. The variable INDGRSAL is derived by estimating the growth in percent

of industry sales between 1981 and 1985, as reported in Trinet.

INDCONC. Industry concentration has in standard IO theory been argued to have a positive

relationship with industry profitability (Bain, 1956; Porter, 1980). The reason being that scale

economies and other sources of market power reduces the threat from potential entrants. In

addition, the risk that incumbents will undertake retaliatory actions against an entrant

increases with industry concentration. This leads us to expect a negative relationship between

industry concentration and the probability of entry. The variable INDCONC is derived by

estimating the four-firm concentration ratio of each industry for 1981, based on Trinet Data.

INDPROF. The relationship between entry and industry profitability is uncertain. On the one hand an industry can only sustain high profitability if entry barriers are high (Baumol et al., 1982), on the other hand these high levels of profitability will be attractive to potential entrants. The net effect of industry profitability on the probability of entry is therefore uncertain. To control for industry profitability we calculated a measure of the median return on assets for each industry over the period 1980 – 1982. The procedure used here calls for some elaboration. The Compustat database which was used to derive INDPROF consists of a segment database which report ROA in four-digit SIC codes, and a corporate database which report ROA on the firm level. We used all observations in the segment database and all single business firms in the corporate database to calculate INDPROF. However this creates a problem because the ROA measures are not directly comparable over these two databases. Because of incomplete asset allocation, ROA is systematically higher in the segment database. To preserve observations we calculated the mean of all observations in each database for each year, and we subsequently divided each individual observation by this mean. Thus, the individual observations were standardized as deviations in percent from the database mean for the relevant year. This allowed us to use observations from both databases and all three years, and we subsequently calculated the median of this measure for each industry. Where a minimum of five observations were obtained, this was done on the level of four digit SIC-industries, if less than five observations were obtained the same measure was calculated on the three digit level, if still less than five observations where obtained, the measure was calculated on the two digit level (following Berger and Ofek, 1995). We expect a negative relationship between INDPROF and the probability of exit.

Firm Level Independent Variables

PARSIZE. The size of the parent has been used as an indicator of a parent's level of financial-, managerial- and other resources. Thus it may be the case that for any given level of relatedness, a large firm is more willing and able to undertake an entry attempt. We try to control for this by including parent size as a control variable. We expect a positive relationship between parent size and the probability of entry. The variable PARSIZE is measured as the total sales of the parent in 1981, based on Trinet data.

PARDIV. The number of SIC-codes a parent is active in may also be a signal of the strategy and the motives of the parent. A parent active in numerous SIC-codes may be following a strategy of broad or unrelated diversification, which makes it more likely that it will enter additional unrelated industries in the future. This could for example be because the firm is dominated by managerial motives of growth or excessive risk reduction, or because decision makers believe that the firm possesses a special ability to create value from general resources, and/or handle a complex organization. Thus we expect a positive relationship between PARDIV and the probability of entry for any level of relatedness. The variable PARDIV is measured as the number of SIC-codes participated in by the parent in 1981, based on Trinet data.

Relatedness Variables

The survivor-based relatedness measures employed in the test of H1-H3 have been presented above. The measures used in testing H1 are called $SURVTOT_i$ and $SICTOT_i$ and capture the sales weighted average relatedness of the target industry *i* to *all other* businesses in the parent *k*. We expect $SURVTOT_i$ to perform better than $SICTOT_i$. Note that the prefix "SURV" indicates a survivor-based measure, while the prefix "SIC" indicates a SIC-based measure.

The measures used in testing H2 are called $SURVNBOR_i$ and $SICNBOR_i$ and capture the weighted relatedness between the target industry i and the two closest neighboring businesses of the parent. We expect $SURVNBOR_i$ to perform better than $SICNBOR_i$. In testing H3 we included all four of these measures, and we hypothesize that the two measures with the prefix "SURV" will perform better than both measures with the prefix "SIC".

Dependent Variable

The dependent variable used to test H1 and H2 is dichotomous. If a parent entered a four-digit SIC code it was not present in in 1981 by 1985, the dependent variable is given a value of 1. If the parent did not enter the industry in question, the value assigned is 0. The Trinet database was used to identify entries and nonentries.

Variable definitions, datasources and predicted signs are summarized in Table 1 below, while Table 2 shows the means, standard deviations and correlation coefficients for all independent variables.

Insert table 1 and 2 about here

----

6. Results

The results from the logistic regression analyses are presented in table 3. Table 3 contains five different logistic regression models. Model 1 contains control variables only. Model 2 contains control variables plus the survivor-based measure $SURVTOT_i$. Model 3 contains control variables plus the equivalent SIC-based measure $SICTOT_i$. Model 4 contains control variables plus the survivor-based measure $SURVNBOR_i$. Model 5 contains control variables plus the equivalent SIC-based measure $SICNBOR_i$.

Insert table 3 about here

As shown in table 3 all four relatedness measures are positively signed and significant at the 0.001 level. This, and the substantial increases in all measures of model performance when any measures of relatedness is included, provides strong and consistent support for the general hypothesis that relatedness matters for patterns of diversification.

H1 predicted that model 2 would explain the probability of exit significantly better than model 3. This hypothesis is strongly supported. All measures of model performance improve substantially when the measure $SURVTOT_i$ is substituted for $SICTOT_i$. The model chi-square improves from 227.26 to 744.1 an improvement which is significant at the 0.001 level. The two pseudo r^2 measures increase from 9.1 % to 26.9 % (Cox and Snell r^2) and from 12.2 % to 35.8 % (Nagelkerke r^2). These increases represent improvements of approximately 195 %. In addition the ability to correctly predict entries and non-entries increased from 61.4 % to 74.4%. This represents an improvement in prediction of 119.2 %. We interpret these results as strong support for H1.

123

H2 predicted that model 4 would explain the probability of exit significantly better than model 5. This hypothesis is also strongly supported. Again, all measures of model performance improve substantially when the measure $SURVNBOR_i$ is substituted for $SICNBOR_i$. The model chi-square improves from 446.38 to 714.60, an increase which is significant at the 0.001 level. The two pseudo r^2 measures increase from 17.1 % to 26.0 % (Cox and Snell r^2) and from 22.8 % to 34.6 % (Nagelkerke r^2). These increases represent improvements of approximately 52 %. Furthermore the ability to correctly predict entries and non-entries increases from 68.2 % to 74.7 %, representing an improvement of 37 %. We interpret these findings as strong support for H2.

H3 predicted that both survivor-based measures would perform significantly better than both SIC-based measures. This hypothesis is supported if the lowest performing survivor-based measure performs significantly better than the best performing SIC-based measure. As shown in table 3 the lowest performing survivor-based measure is $SURVNBOR_i$ while the best performing SIC-based measure is $SICNBOR_i$. The test of H3 therefore reduces to the same test as H2, which as noted was strongly supported. We therefore conclude that H3 is also strongly supported.

Another noteworthy observation is the difference in performance between the two SIC-based measures. $SICNBOR_i$ outperformed $SICTOT_i$ substantially, as can be seen by a difference in chi-square between model 3 and 5 of 219.12 which is significant at the 0.001 level. The pseudo r^2 measures increase from 9.1 % to 17.1 % (Cox and Snell r^2) and from 12.2 % to 22.8 % (Nagelkerke) as we move from model 3 to model 5. A possible reason for this finding is the coarse grained nature of the SIC distance measures (only three different levels of "distance"). This means that every distance between the target industry and an existing business will be likely to include a substantial measurement error, due to this coarse grained scale. The $SICTOT_i$ measure may therefore imply multiplying more noise into the measure than

SURVNBOR$_i$. simply because the number of distances included is higher when the target

industry is matched to all businesses in the parent portfolio (and not only the two closest

neighbors).

In contrast. among the survivor-based measures SURVTOT$_i$ outperformed SURVNBOR$_i$.

The difference in chi-squares among model 2 and 4 is 29.5, which is significant at the 0.01

level, but the differences in the pseudo r^2 measures are quite small. One possible explanation

for the apparent superiority of SURVTOT$_i$ over SURVNBOR$_i$ is that the continuous scale

underlying the survivor-based measures means that the problem of multiplying more noise

into the measure is substantially reduced. and that the measure SURVTOT$_i$ benefits from

being able to capture relatedness effects that extend beyond the two closest related industries.

7. Conclusions and Caveats

In sum. our findings are that the survivor-based measures outperform equivalent SIC-based

measures in terms of explaining the probability of entry. Our findings also indicate support for

the hypothesis that relatedness matter for patterns of diversification, since our data shows that

relatedness significantly increases the probability of entry. However, in drawing these

conclusions some caution is warranted.

One reason for caution is that entry decisions may be influenced by other factors than

relatedness that we have not satisfactorily controlled for. The entry decisions observed here

may for example be more influenced by fashion and herd behavior than conscious attempts to

exploit relatedness (i.e. it is fashionable to enter "similar" businesses). This is both a threat to

interpreting our findings as support for the relatedness hypothesis, but more importantly, it

may be the case that the survivor-based measures perform better than the SIC-based measures

because they capture such herd behavior better. This suspicion arises because the survivor-

based measures are based on observing the combinations chosen by other firms in the same industries in a period recently preceding the decision period. These "neighboring" firms are presumably the ones the focal firm would be herding after. The data supplied here can not rule out this possibility.

Another alternative interpretation is that the survivor-based measures outperform the SIC based measures because of contamination from motives associated with multipoint competition and mutual forbearance. This refers to a tendency among firms competing across several markets to instigate a balance of terror where competition is less aggressive than it would have been otherwise (Edwards, 1955; Gimeno, 1999; Greve and Baum, 2001). Firms may prefer to enter industries where they will meet existing competitors as a mean towards establishing mutual forbearance. The way the survivor-based measures are constructed makes them particularly well suited to capture such motives, since a related industry - measured this way - would be an industry where the firm will meet many of its existing competitors. Future work should therefore emphasize testing both the herd behavior and the mutual forbearance interpretation of the findings in this study, as well as testing the survivor-based measures against other dependent variables which relatedness is supposed to affect. In particular it would be useful to test the survivor-based measures against measures of firm level performance, such as ROA, tobins-q and firm growth. And surely it would also be desirable with replications using data from other periods and other places.

These important caveats not withstanding, it is our interpretation that the present findings, particularly in combination with the results reported by Lien (2003) on exit data, strongly suggests that survivor-based measures can make a valuable contribution to research on how relatedness affects other variables of interest (i.e. what relatedness does). As a minimum it would seem that the results in these two papers justify additional work on validating the survivor-based approach to measuring relatedness.

Refrences

Alchian AA. 1950. Uncertainty, Evolution, and Economic Theory. *Journal of Political Economy* **LVIII**: 211-221.

Amihud Y, Lev B. 1981. Risk Reduction as a Managerial Motive for Conglomerate Mergers. *Bell Journal of Economics* **12**: 605-617.

Bain JS. 1956. *Barriers to new competition, their character and consequences in manufacturing industries*. Harvard University Press: Cambridge.

Baumol WJ, Panzar JC, Willig RD. 1982. *Contestable markets and the theory of industry structure*. Harcourt Brace Jovanovich: New York.

Berger PG, Ofek E. 1995. Diversifications Effect on Firm Value. *Journal of Financial Economics* **37**(1): 39-65.

Caves RE, Porter ME, Spence AM. 1980. *Competition in the open economy : a model applied to Canada*. Harvard University Press: Cambridge, Mass.

Chatterjee S, Wernerfelt B. 1991. The Link between Resources and Type of Diversification - Theory and Evidence. *Strategic Management Journal* **12**(1): 33-48.

Chatterjee S, Blocher JD. 1992. Measurement of Firm Diversification - Is It Robust. *Academy of Management Journal* **35**(4): 874-888.

Christensen JF, Foss NJ. 1997. Dynamic corporate coherence and competence based competition: Theoretical foundations and strategic implications. In A Heene, R Sanchez (Eds.), *Competence-based Strategic Management*. Wiley: Chichester.

Edwards CD. 1955. Conglomerate bigness as a source of power, *NBER conference report Business Concentration and Economic Policy*. Princeton University Press.: Princeton.

Elster J. 1989. *Nuts and Bolts for the Social Sciences*. Cambridge University Press: New York.

Fan JPH, Lang LHP. 2000. The measurement of relatedness: An application to corporate diversification. *Journal of Business* **73**(4): 629-660.

Farjoun M. 1994. Beyond Industry Boundaries - Human Expertise, Diversification and Resource-Related Industry Groups. *Organization Science* **5**(2): 185-199.

Foss NJ, Christensen JF. 2001. A market-process approach to corporate coherence. *Managerial and Decision Economics* **22**: 213-226.

Friedman M. 1953. The Methodology of Positive Economics, *Essays in Positive Economics*: 328. University of Chicago Press: Chicago.

Gimeno J. 1999. Reciprocal threats in multimarket rivalry: Staking out 'spheres of influence' in the US airline industry. *Strategic Management Journal* **20**(2): 101-128.

Greve HR, Baum JAC. 2001. Introduction: A multiunit, multimarket world, *Advances In Strategic Management*, Vol. 18: 1-28.

Hodgson G. 1993. *Economics and Evolution*. University of Chicago Press: Ann Arbor.

Hoskisson RE, Hitt MA, Johnson RA, Moesel DD. 1993. Construct Validity of an Objective (entropy) Categorical Measure of Diversification. *Strategic management Journal* **14**: 215-235.

Jaquemin A, Berry C. 1979. Entropy Measure of Diversification and Corporate Growth. *Journal of Industrial Economics* **27**: 359-379.

Jensen MC, Meckling WH. 1976. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics* **3**(4): 305-360.

Klein PG, Klein SK. 2001. Do entrepreneurs make predictable mistakes? Evidence from corporate divestures. *The Quarterly Journal of Austrian Economics* 4: 3-23.

Kwoka JE, Ravenscraft DJ. 1986. Cooperation V Rivalry - Price-Cost Margins by Line of Business. *Economica* 53(211): 351-363.

Lien. 2003. Yet another way of measuring relatedness - this one: Let competition do it! *This thesis: paper no. 2.*

Markides CC, Williamson PJ. 1994. Related Diversification, Core Competence and Corporate Performance. *Strategic Management Journal* 15: 149-165.

Markides CC, Williamson PJ. 1996. Corporate Diversification and Organizational Structure: A Resource-Based View. *Academy of Management Journal* 39: 340-367.

McFadden D, Manski CF. 1981. Alternative estimators and sample designs for discrete choice analysis. In D McFadden, CF Manski (Eds.), *Structural analysis of discrete data with econometric applications:* xxv, 477 s. MIT Press: Cambridge, Mass.

Milgrom P, Roberts J. 1992. *Economics, organization and management.* Prentice-Hall: Englewood Cliffs, N.J.

Montgomery CA, Hariharan S. 1991. Diversified Expansion by Large Established Firms. *Journal of Economic Behavior & Organization* 15(1): 71-89.

Montgomery CA. 1994. Corporate Diversification. *Journal of Economic Perspectives* 8(3): 163-178.

Mueller DC. 1969. Theory of Conglomerate Mergers. *Quarterly Journal of Economics* 83(4): 643-659.

Penrose E. 1959. *The theory of the growth of the firm.* Blackwell: Oxford

Porter ME. 1980. *Competitive strategy : techniques for analyzing industries and competitors.* Free Press: New York.

Richardson GB. 1972. The Organisation of Industry. *Economic Journal* 82(327): 883-896.

Robins J, Wiersema MF. 1995. A Resource-Based Approach to the Multibusiness Firm - Empirical-Analysis of Portfolio Interrelationships and Corporate Financial Performance. *Strategic Management Journal* 16(4): 277-299.

Robins JA, Wiersema MF. 2003. The measurement of corporate portfolio strategy: Analysis of the content validity of related diversification indexes. *Strategic Management Journal* 24(1): 39-59.

Rumelt RP. 1974. *Strategy, structure, and economic performance.* Division of Research Graduate School of Business Administration Harvard University ; distributed by Harvard University Press: Boston. Cambridge, Mass.

Rumelt RP. 1982. Diversification Strategy and Profitability. *Strategic Management Journal* 3(4): 359-369

Salinger MA. 1984. Tobin-Q, Unionization, and the Concentration-Profits Relationship. *Rand Journal of Economics* 15(2): 159-170

Schmalensee R. 1989. Inter-Industry Studies of Structure and Performance. In R Schmalensee, RD Willig (Eds.), *Handbook of Industrial Organization.* North Holland: New York

Sharma A. 1998. Mode of Entry and Ex - Post Performance. *Strategic Management Journal* 19: 879-900

Shleifer A, Vishny RW. 1994. Takeovers in the 1960s and the 1980s: Evidence and Implications. In RP Rumelt, DE Schendel, DJ Teece (Eds.), *Fundamental Issues in Strategy.* Harvard Business School Press: Boston.

Silverman BS. 1999. Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. *Management Science* 45(8): 1109-1124.

Teece DJ. 1980. Economies of Scope and the Scope of the Enterprise. *Journal of Economic Behavior & Organization* 1(3): 223-247.

Teece DJ. 1982. Towards an Economic-Theory of the Multiproduct Firm. *Journal of Economic Behavior & Organization* 3(1): 39-63.

Teece DJ, Rumelt R, Dosi G, Winter S. 1994. Understanding Corporate Coherence - Theory and Evidence. *Journal of Economic Behavior & Organization* 23(1): 1-30.

Voight CW. 1993. The Trinet large establishment data. University of california at Los Angeles, CA: Anderson Graduate School of Management.

Williamson OE. 1985. *The economic institutions of capitalism : firms, markets, relational contracting.* Free Press: New York.

Winter SG. 1971. Satisficing, Selection, and Innovating Remnant. *Quarterly Journal of Economics* 85(2): 237-261.

Table 1:   Variable definitions, datasources and predicted signs

Variable	Definition	Data Source	Predicted Sign
INDGRSAL	Salesgrowth in industry i between 1981 and 1985	Trinet	+
INDCONC	4-firm concentration ratio in 1981 in industry i	Trinet	-
INDPROF	Industry median ROA 1980-1982 in industry i	Compustat	?
PARSIZE	Total sales of the parent in 1981	Trinet	+
PARDIV	Number of 4-digit SIC codes participated in by parent in 1981	Trinet	+
SURVTOT	Weighted average *survivor-based* relatedness of industry i to all industries in the portfolio of the parent	Trinet	+
SICTOT	Weighted average *SIC-* based relatedness of industry i to all industries in the portfolio of the parent	Trinet	+
SURVNBOR	Weighted *survivor-based* relatedness of industry i to the two closest businesses of the parent	Trinet	+
SICNBOR	Wheighted *SIC*-based relatedness of industry i to the two closest businesses of the parent	Trinet	+

Table 2   Means, Standard Deviations and Correlation Coefficients of Independent Variables   N=2378

	Variables	Mean	SD	Indgrsal	Indconc	Indprof	Parsize	Pardiv	Survtot	Sictot	Survnbor
	Industry variables										
1	Indgrsal	0,602	1,644								
2	Indconc	32,21	21,42	0,15***							
3	Indprof	0,093	0,455	0,01	-0,04***						
	Firm variables										
5	Parsize	48727	56790	0,00	0,05***	-0,04**					
6	Pardiv	39,81	22,55	-0,02	0,06***	-0,05***	0,38***				
	Relatedness variables										
8	Survtot	3,41	4,43	0,00	-0,06***	0,04**	-0,08***	-0,09***			
9	Sictot	0,069	0,181	-0,01	0,02	0,08***	-0,07***	-0,08***	0,44***		
10	Survnbor	12,75	9,39	0,01	-0,09***	0,02	0,04**	0,15***	0,67***	0,29***	
11	Sicnbor	0,630	0,719	0,02	-0,05***	0,04*	-0,06***	0,23***	0,37***	0,47***	0,51***

*** Correlation is significant at the 0,01 level
** Correlation is significant at the 0,05 level
* Correlation is significant at the 0,1 level

Table 3: Logistic Regression Output   N = 2378

	Model 1 Ctrls Only		Model 2 Ctrls + Survtot		Model 3 Ctrls + Sictot		Model 4 Ctrls + Survnbor		Model 5 Ctrls + Sicnbor		Signs	
	Beta	Wald	Beta	Wald	Beta	Wald	Beta	Wald	Beta	Wald	Pred	Obs
Constant	0,39***	13,82	-0,84**	41,71	0,16*	2,17	-1,07***	64,77	0,00	0,00		
Indgrsal	0,05**	3,99	0,06***	4,06	0,06**	4,90	0,04	1,72	0,04*	2,68	+	+
Indconc	-0,01***	47,65	-0,01***	31,88	-0,02***	53,51	-0,01***	27,95	-0,01***	37,75	-	-
Indprof	-0,00	0,00	-0,13	1,43	-0,10	1,21	-0,07	0,44	-0,09	0,81	?	
Parsize	0,00	0,00	0,00**	4,85	0,00	0,63	0,00	0,63	0,00	0,34	+	
Pardiv	0,00	0,38	0,00	0,33	0,00	0,64	-0,01***	21,03	-0,01***	17,07	+	
SURVTOT$_i$			0,40***	403,33							+	+
SICTOT$_i$					4,40***	99,94					+	+
SURVNBOR$_i$							0,16***	427,67			+	+
SICNBOR$_i$									1,30***	326,54	+	+
-2Log likelihood	3245,76		2552,23		3069,06		2581,73		2849,95			
ModelChi-square	50,56***		744,09***		227,26***		714,60***		446,38***			
Δ Chi-square vs. model 1			693,53***		176,70***		664,03***		395,82***			
Δ Chi-square model 2 vs. model 3					516,83***							
Δ Chi-square model 4 vs. model 5									268,22***			
Cox and Snell r^2	0,021		0,269		0,091		0,260		0,171			
Nagelkerke r^2	0,028		0,358		0,122		0,346		0,228			
Correct classifications	55,3%		74,4%		61,4%		74,7%		68,2%			

*** Correlation is significant at the 0,01 level
** Correlation is significant at the 0,05 level
* Correlation is significant at the 0,1 level

132

Survivor-based Measures of Relatedness: Two Alternative Interpretations

Lasse B. Lien

Institutt for strategi og ledelse

Norges Handelshøyskole

Breiviksveien 40

N-5045 Bergen

Norway

lasse.lien@nhh.no

Paper no. 4

Survivor-based Measures of Relatedness: Two Alternative Interpretations

Abstract

Recently a survivor-based approach to measuring relatedness has been suggested as superior to the traditional SIC-based measures. Empirical studies have found survivor-based measures to be superior in terms of predicting both entry and exit decisions by diversified firms. However, the existing evidence cannot rule out two rival interpretations. One is that herd behavior constitutes an important influence on decision makers, and that the survivor-based approach captures such unwanted influences better than SIC-based measures. Survivor-based measures are indeed constructed from data about "what others do", and the apparent superiority of the survivor-based measures may be an artifact of their ability to capture herding tendencies. Another alternative interpretation is associated with the motive of mutual forbearance through multi-point competition. The survivor-based measures are also particularly well suited to capture such motives, because relatedness measured this way will be strongly and positively correlated with multi-market contact. This paper examines whether these "non-relatedness" explanations may account for the alleged superiority of the survivor-based measures over SIC-based alternatives.

1. Introduction

In two recent papers (Lien, 2003 a; Lien, 2003 b) a survivor-based approach to relatedness was found to be superior to conventional SIC-based measures in terms of explaining entry and exit decisions made by diversified firms. This paper explores the possibility that these findings do not result from a better ability to capture relatedness, but from mechanisms and motives that are unrelated to the relatedness concept. In particular there are two plausible alternative hypotheses.

The first is associated with the impact of herd behavior (Banerjee, 1992; Scharfstein and Stein, 1990). Given that the survivor-based approach to relatedness is built on the idea that what is related is what firms combine often, such a measure is susceptible to falsely interpret what is essentially mimetic- or herd behavior as an indication of (or a response to) true relatedness. SIC-based measures, on the other hand, are not influenced by the actions of others, and they are therefore less likely to be contaminated by such mechanisms.[40] In short, the observed superiority of the survivor-based measures in terms of predicting entry and exit decisions may be an artifact of their ability to capture mimetic tendencies, and not due to a superior ability to pin down relatedness.

Secondly, the survivor-based measures may be contaminated by motives related to market power. There is a substantial literature suggesting a relationship between multi market contact and mutual forbearance. This refers to a tendency among firms competing across several markets to instigate a balance of terror where competition is less aggressive than it would have been otherwise (Edwards, 1955; Gimeno, 1999; Greve and Baum, 2001). Again, the way the survivor-based measures are constructed makes them particularly well suited to capture

---

[40] The SIC-based measures are based on standardized distances in the SIC-system.

such motives, since they are indeed built from a count of frequencies of multi market contact. One may therefore speculate that the superior ability of the survivor-based approach to explain patterns of entry and exit is not due to relatedness, but to a particular sensitivity to the motives of creating and exploiting benefits from mutual forbearance. SIC-based measures are not equally sensitive to these motives, because they are not constructed on the basis of multi market contact.

This paper examines these two alternative interpretations of the findings reported by Lien (2003a, 2003b). The paper is organized as follows: Section 2 presents the logic underlying the survivor-based measures, and how different survivor-based measures are derived from this logic. Section 3 presents the herd behavior mechanism, and develops hypotheses that test whether herd behavior can explain the apparent superiority of the survivor-based measures. Section 4 is an empirical test of this herd behavior reinterpretation. Section 5 presents the mutual forbearance motive, and develops hypotheses that test whether this can account for the apparent superiority of the survivor-based measures. Section 6 is an empirical test of this mutual forbearance reinterpretation. Section 7 concludes.

2.   The Survivor-based Approach to Relatedness

2.1   The Logic

The fundamental premise of the survivor-based approach to relatedness is that industries that are related will be more frequently combined inside firms. The term "more frequently" is operationalized by estimating how much the frequencies of *actual* combinations of four-digit SIC industries inside firms deviate from what one would expect if diversification patterns were *random* (Teece, Rumelt, Dosi and Winter, 1994). A given pair of industries is thus related if the actual number of firms combining this pair of industries exceeds the number one

would expect if firms combined industries randomly. If this difference is large and positive they are closely related, if it is zero or negative they are unrelated.

The potential advantages of using this survivor-based approach stem from two main sources. One is that it better incorporates the knowledge of the best informed actors which presumably are those making portfolio decisions. Since the survivor-based approach is built from observing the actions of those holding superior information, one may stipulate that more and better information is reflected in these measures (i.e. a Hayekian argument). Secondly, and equally important, even if the information of these decision makers is poor, and/or their motives are illegitimate, their decisions have been screened by the competitive process (Alchian, 1950). If the screening process of competitive markets is efficient, poor decisions will either be reversed, or the firm that made them will loose resources and ultimately become extinct. This second argument relies crucially on what has been termed the survivor principle (hence the name "survivor-based measures"). The survivor principle has as it core idea the notion that the competitive process screens for efficiency - and does so well enough that a sample of competitive firms will be dominated by the decisions or behaviors that are efficient - at least in a comparative sense (Alchian, 1950:211).[41] In this particular setting this insight is applied to portfolio decisions, and it is assumed that the competitive process is quite effective in removing inefficient combinations.[42]

In sum, the potential advantages of a survivor-based approach stems from better use of the information revealed by the actions of those holding superior information, and the screening of these decisions by the competitive process.

2.2   The Survivor-based Measures

---

[41] For the view that the competitive process creates outcomes that are optimizing, cfr. Friedman (1953).
[42] For critique of this optimistic view of the screening ability of the competitive process cfr. Hodgson (1993), Winter (1971), Elster (1989).

All versions of the survivor-based approach relies fundamentally on estimates of how much the frequencies of *actual* combinations of four-digit SIC industries inside firms deviate from what one would expect if diversification patterns were *random*. We shall first show how this key number is computed, before we turn to how it can be used to derive different measures of survivor-based relatedness. The following draws heavily on Teece, Rumelt, Dosi and Winter (1994).

Let the universe of diversified firms consist of K firms, each active in two or more of I industries. Let $C_{ik} = 1$ if firm k is active in industry i. The number of industries participated in by firm k is $m_k = \Sigma_i C_{ik}$ and the number of diversified firms present in industry i is $n_i = \Sigma_k C_{ik}$. Let J_{ij} be the number of diversified firms active in both industries i and j, such that

$J_{ij} = \Sigma_k C_{ik}C_{jk}$. Thus J_{ij} is a count of how often industries i and j are actually combined within the same firm. J_{ij} will be larger if industries i and j are related, but will also increase with n_i and n_j. To remove the effect of the size of industries i and j, the number J_{ij} is compared with the number of expected combinations if diversification patterns were random.

The random diversification hypothesis can be operationalized as a hypergeometric situation where a sample of size n_i is drawn (without replacement) from a population of K firms. Those chosen are considered active in industry i. A second independent sample of size n_j is then drawn from the population the population of K firms. Those chosen are considered active in industry j. The number x_{ij} of firms active in both i and j is then a hypergeometric random variable with population K, special members n_i and sample size n_j. The distribution function for this variable is then:

$$\Pr(X_{ij} = x) = f_{hg}(x, K, n_i, n_j) = \frac{\binom{n_i}{x}\binom{K - n_i}{n_j - x}}{\binom{K}{n_j}}$$

The mean and variance of X_{ij} are:

$$\mu_{ij} = E(X_{ij}) = \frac{n_i n_j}{K},$$

$$\sigma^2 = \mu_{ij}\left(1 - \frac{n_i}{K}\right)\left(\frac{K}{K-1}\right).$$

A standardized measure of the relatedness between industries i and j is then constructed based on the difference between J_{ij} and μ_{ij} in the following fashion:

$$SR_{ij} = \frac{J_{ij} - \mu_{ij}}{\sigma_{ij}}$$

The measure SR_{ij} is thus a standardized measure of how much actual combinations exceed expected combinations under the random diversification hypothesis. With this fundamental measure in hand it is possible to compute several survivor-based measures of how related a given business (or a potential destination industry) is to the businesses in a given parent portfolio. We present the three measures used by Lien (2003a, 2003b).

The first measure captures the weighted average relatedness of a business i (or a destination industry i[43]) to all other businesses in the portfolio of a given parent. Assume a diversified firm that participates in m industries. Its business in industry j has sales of s_j and survivor-based relatedness SR_{ij} with industry i. The weighted average relatedness SURVTOT$_i$ of the business in industry i to all other businesses in the firm is then defined as:

$$SURVTOT_i = \frac{\sum_{j \neq i} SR_{ij} s_j}{\sum_{j \neq i} s_j}$$

The second measure does not consider how related a given business i is to *all* other business in the corporate portfolio, but how related it is to the *two closest* neighboring businesses of the

---

[43] In the following the term "business i" is substitutable with destination "industry i". The latter expression is appropriate for analyses of entry decisions, while the former is appropriate for exit decisions.

parent. The approach here is to rank the survivor-based measure SR_{ij} between the business i, and all other businesses in the parent portfolio. The two industries with the highest measure of SR_{ij} are considered the neighboring businesses. Let $\lambda_{ij} = 1$ for a business that is defined as a neighbor to business i, and $\lambda_{ij} = 0$ for those that are not. The weighted average relatedness of neighbors to business i is then defined by:

$$SURVNBOR_i = \frac{\sum_{j \neq i} SR_{ij} s_j \lambda_{ij}}{\sum_{j \neq i} s_j \lambda_{ij}}$$

The third measure captures how related a business i is to the parent's core business (measured as the largest 4-digit SIC code in terms of revenue). Let SR_{ic} be the survivor-based measure between a given industry i, and the core business c. This measure includes consideration of the size of the focal business (where size refers to share of the parent's total sales), and construct it so that its value is high for large businesses close to the core business, and smaller for a small businesses distant to the core business. This measure $SURVCORE_i$ is defined as:

$$SURVCORE_i = SR_{ic} \frac{s_i}{\sum_i s_i}$$

2.3   Preceding Findings

To validate these survivor-based measures of relatedness Lien (2003a, 2003b) compared them with parallel SIC-based measures[44] in terms of their ability to predict entry and exit decisions by diversified firms. In Lien (2003b) the measures SURVTOT and SURNBOR were compared with SIC-based equivalents with respect to prediction of *entry* decisions. Based on relatedness measures from 1981, the ability to predict whether entry would occur between 1981 and 1985 was estimated. In Lien (2003a) the measures SURVTOT and SURVCORE were compared with SIC based equivalents with respect to prediction of *exit* decisions. Based

---

[44] The procedures for calculating SIC-based equivalents to the survivor-based measures are described in section 4.1. In general this is done by substituting distances in the SIC-system for the survivor-based measure SR_{ij} in the formulas provided above.

on relatedness measures from 1981 the ability to predict which businesses a diversified firm had chosen to exit by 1985 was estimated.

In both cases (entry and exit) the survivor-based measures were found to score significantly higher than their SIC- based equivalents on all measures of model performance. Not only did each survivor-based measure outperform its SIC-based equivalent, but all survivor-based measures were found to outperform all SIC-based measures.

While these findings appear impressive they may be so for the wrong reasons, meaning that they are not necessarily the result of a superior ability to measure relatedness. We now turn to the first of two alternative interpretations of these findings.

3.   A Herd Behavior Interpretation

3.1 The Logic of Herd Behavior

John Maynard Keynes was the first to suggest that herd behavior may play a role in economic life. In *The General Theory*, Keynes (1936, pp. 157-58) noted the following in a passage about the irrational behavior of investors:

> *"Wordly wisdom teaches that it is better for reputation to fail conventionally than to succeed unconventionally"*

What Keynes suggests here is that managers concerned about their reputation may choose to follow the herd, even when their private information tells them that they should not. This insight has been formalized by Scharfstein and Stein (1990) who demonstrate that managers may indeed be reluctant to act according to his/her own information and beliefs, because a bad decision is not as damaging for a manager's reputation when others make the same

mistake. This incentive bias may produce aggregate outcomes where private information is not efficiently used, and where decision makers converge on decisions that are not efficient.

Banerjee (1992) presents a slightly different version. He investigates a situation where each decision maker pays heed to what other decision makers are doing, not to protect his/her reputation with a principal or in managerial labor markets, but because the actions of others may reflect private information that the others have and the focal decision maker does not. This, however, implies that each decision maker becomes less responsive to his/her own information, and accordingly the actions of each decision maker become less informative to others. The outcome may again be an equilibrium where private information is not efficiently used and decision makers converge on decisions that are not efficient.

3.2 Relatedness Measures and Herd Behavior

With respect to patterns of diversification the herding mechanisms presented above implies that decisions about which industries to combine inside a firm can be substantially influenced by the actions of others, and not superior local knowledge about which businesses are related to which. For example entry into a new industry may be considered less risky in terms of managerial reputation if this creates combinations that have been chosen by many others. Or alternatively, if a particular combination of industries has been chosen by many others, a manager may believe that these have private information about the benefits of combining it with one of the existing businesses.

What is important to note here is that SIC-based measures and survivor-based measures are likely to differ with respect to contamination by herd behavior. The reason is simply that the survivor-based approach involves measuring which combinations firms in the same industries as the focal firm have chosen in a period recently preceding the decision period (cfr the calculations of SR_{ij} in section 2.2). These are presumably the firms the focal firm would be

herding after. The SIC-based approach, on the other hand, is based on standardized distances in the SIC-system, and it plainly does not reflect what "others are doing" nearly as well. There are for example numerous examples of industries that are close in the SIC-system, but are never combined inside firms, and the reverse; that firms are distant in the SIC-system, but frequently combined.

Lien (2003b) assumes that decision makers make decisions independently, and that the propensity to enter common combinations is a result of decision makers facing similar facts (about relatedness). However, the decision to follow suit may be driven by the herding mechanisms just described, and not efficient use of available information. The relative success of the survivor-based measures in terms of predicting patterns of entry reported by Lien (2003b) may therefore arise from a superior ability to capture herd behavior and not relatedness.

If this is the case, we would expect the survivor-based relatedness measures to display superior performance in terms of predicting entry decisions, as Lien actually found, but not to be superior in terms of predicting post entry performance. This means that while herd behavior may influence entry decisions, once entry has occurred competitive forces and economic reality sets in and begins its work of screening the good decisions from the bad. The worse the post entry performance, the more likely that the entered business is exited again (i.e. the entry decision is reversed). If SIC- and survivor-based measures capture true relatedness equally well, we would expect no difference between the two approaches in terms of predicting which of the entry decision that will subsequently be reversed. Based on this logic we may formulate the following hypotheses.

H1:   Survivor-based related measures are superior to SIC-based measures in terms of predicting entry decisions

H2:   Survivor-based relatedness measures are not superior to SIC-based measures in terms of predicting post entry survival

H1 was supported in Lien (2003b), but his interpretation did not include the herding mechanisms discussed above. We shall therefore focus on testing H2, and we suggest that support for H2 implies support for a herd behavior interpretation of the reported findings regarding H1. A contradiction of H2 can be considered support for the suggestion that survivor-based measures are better able to capture relatedness than SIC-based measures, since it means it will have withstood an attempt at falsification based on herd behavior mechanisms.

4.   Testing the Herd Behavior Interpretation

4.1 Data and Methodology

Sample

The data used to test H2 is a sub-sample of the sample used by Lien (2003b).[45] The full sample used by Lien consisted of 1202 instances of entry and 1176 instances of non-entry occurring between 1981 and 1985. Ideally we could have used all the 1202 cases of entry for the purpose of the present work, but this was not possible for two reasons. One is that 1987 was the latest date for which we could obtain usable data regarding industry participation, and hence could register whether a firm was still active in an entered industry. Given that we wanted some time for the competitive process to work on the entry decisions, we felt it was necessary to restrict the sample to entries made between 1981 and 1983. This reduced the

---

[45] For a more detailed description of this sample, cfr. Lien (2003b)

sample to 401 cases of entry. Using data from 1987 to measure post entry survival also posed an additional challenge. The SIC-system was revised in 1987, and there is no clear cut procedure for conversion between the two versions of the SIC-system. Fortunately, of the 913 relevant four digit SIC-codes,[46] a total of 609 SIC codes went unchanged through this revision. To avoid the possibility that the SIC-revision creates confusion regarding whether an entry decision had been reversed, we further restricted our sample to entries into the 609 industries that were not affected by the revision. This further reduced our sample to 229 entries, all occurring between 1981 and 1983, for which their continued existence in 1987 could be unambiguously determined.

Statistical Methods

H2 involves a dichotomous dependent variable, which specifies whether an industry entered between 1981 and 1983 had been exited by 1987. Thus a logistic regression model was considered appropriate. The general model is the following:

$$P(\text{exit}=1) = \beta_1 + \beta_2(\text{Relatedness}) + \varepsilon$$

Note that the purpose of the model is the *relative* performance of different relatedness measures in terms of explaining the probability of an entered business being exited again. It is not to determine the absolute values of the coefficients, nor to explain the probability of exit per se. Including control variables will therefore not shed light on the research question posed here unless they can be expected to correlate differently with the different relatedness measures we are comparing. The only variable we know that may be theoretically expected to do so is industry concentration. However this possibility will be explored in sections 5 and 6

---

[46] SIC-codes referring to public and non-profit industries are omitted.

of this paper. We therefore omit control variables in the present study, but we note that the findings reported below did not change when controls for growth, industry profitability, parent size, parent diversity, parent leverage and parent liquidity were included (the regressions with control variables included are available from the author).

Independent Variables

The independent variables used in this study are the four different relatedness measures used in Lien (2003b). The two survivor-based measures are SURVTOT and SURVNBOR, the procedure for calculating these have been presented in section 2.2, but we reiterate that SURVTOT measures the sales weighted average relatedness of the entered industry i to all other industries a parent is active in. SURVNBOR captures the weighted relatedness of an entered industry i to the two closest related industries the parent is active in. The two SIC-based measures parallel these survivor-based measures. SICTOT is an equivalent to SURVTOT based on SIC-distances. SICTOT is calculated as follows:

$$SICTOT_i = \frac{\sum_j d_{ij} s_j}{\sum_j s_j}$$

Where $d_{ij} = 2$ if i and j are in the same 3-digit SIC codes

    $d_{ij} = 1$ if i and j are in different 3-digit-, but the same 2 digit SIC codes

    $d_{ij} = 0$ if i and j are in different 2-digit SIC codes

    $s_j =$ parent sales in industry j

SICNBOR is an equivalent to SURVNBOR based on SIC-distances. SICNBOR is calculated as follows:

$$SICNBOR_i = \frac{\sum_j d_{ij} s_j \lambda_{ij}}{\sum_j s_j \lambda_{ij}}$$

Where $d_{ij} = 2$ if i and j are in the same 3-digit SIC codes

    $d_{ij} = 1$ if i and j are in different 3-digit-, but the same 2-digit SIC codes

$d_{ij} = 0$ if i and j are in different 2-digit SIC codes

$\lambda_{ij} = 1$ if business j is defined as a neighbor to business i [47]

$\lambda_{ij} = 0$ if business j is not defined as a neighbor to business i

Dependent Variable

As already indicated the dependent variable was whether a firm had exited an entered

business by 1987. If yes, a value of 1 was assigned, if not a value of 0 was assigned. Both

divestures and closures were considered to represent exit.

Variable definitions, data source and predicted signs are summarized in Table 1 below. Table

2 shows the means, standard deviations and correlation coefficients for all four independent

variables.

Insert table 1 about here

Insert table 2 about here

4.2   Results and Discussion

The results from the regression analyses are presented in table 3. Table 3 contains four

different logistic regression analyses, one for each relatedness measure. H2 predicted that the

survivor-based measures would *not* outperform SIC-based measures in terms of predicting

---

[47] Note that we resorted to the following procedure when business were equidistant from the target industry: First we identified the closest neighbor, if several where equidistant, we used the sales of the *largest* business to compute sales weights. Then we identified the second closest. If there were several second closest firms, we chose the *smallest* of these. This was done to reflect our assumption that one would expect closer cooperation with the closest neighbor than the second closest. This procedure implies that when the closeness to the two neighbors differed, we weighted our measure in favor of the closest of the two.

post entry survival. As table 3 shows this hypothesis is strongly contradicted. All survivor-based measures perform better than all SIC-based measures in terms of explaining the probability that an entered business will be exited again. In particular we note that the measure SURVTOT perform better than all the other measures, but SURVNBOR also performs significantly better than SICTOT. The difference in chi-square between the models containing these two measures is significant at the 0,01 level. The ability of the measure SICNBOR to predict exit is not statistically significant.

Next we ran a model where both SURVTOT and SICTOT were included. The output from this model is presented in table 4. The results show that when both are included SURVTOT completely replaces the predictive power of SICTOT, and that SICTOT does not add any explanatory power to the model with SURVTOT only.

----

insert table 3 about here

----

insert table 4 about here

These results contradict the suggestion that the superiority of survivor-based measures of relatedness found in Lien (2003b) will disappear when herd behavior is taken into account. They do not, however, indicate that herd behavior does not exist, nor that survivor-based measures are not influenced by such tendencies. What they do suggest is that survivor-based measures *are* superior to SIC-based measures in terms of capturing relatedness, and that this seems to overshadow any noise resulting from inclusion of herding tendencies.

5. A Mutual Forbearance Interpretation

5.1 The Logic of Mutual Forbearance

The mutual forbearance hypothesis is usually credited to Edwards (1955). This hypothesis

suggests that high levels of contact between firms across markets will induce mutual

forbearance, causing multipoint competitors to refrain from aggressively attacking each other.

The incentive to do so stems from the fact that high levels of intermarket contact enable a firm

to respond to an aggressive action by a multipoint rival in markets other than the one in which

the action takes place. This possibility raises the potential costs of aggressive moves and may

therefore induce a state of less vigorous competition than would have ocurred without

multimarket contact (Karnani and Wernerfelt, 1985).

Empirical work has supported the mutual forbearance hypothesis both with respect to patterns

of market entry and competitive outcomes. For example Scott (1989, 1991) has found that

conglomerate firms have higher levels of multimarket contact than would be expected if entry

patterns were independent. Furthermore, multimarket contact has been found to be positively

associated with higher prices (Feinberg, 1985; Evans and Kessides, 1994; Gimeno and Woo,

1996; Gimeno, 1999), higher profits (Barnett et al., 1994; Phillips and Mason, 1996) and

lower exit rates (Barnett, 1993; Baum and Korn, 1996; Baum and Korn, 1999; Baum, 1999).

It seems warranted to conclude that empirical work supports the idea that diversification may

be affected by the possibility of achieving benefits from mutual forbearance. However, it is

unclear how sensitive this strategy is to the existing (and potential) number of firms that are

not multipoint competitors in the market (Greve and Baum, 2001). Under some conditions it

may be rational for such firms to follow the less aggressive path of the forbearing firms (i.e. a

puppy dog/fat cat equilibrium, see Fudenberg and Tirole, 1984; Bulow et al., 1985), while

under other conditions it may be difficult to maintain forbearance due to the threat of entry

(Baumol et al., 1982), or because competition from such incumbents forces multipoint

competitors into vigorous competition (i.e. a prisoner's dilemma equilibrium). But while the

exact sensitivity of the mutual forbearance hypothesis to market structure is uncertain, it does

seem quite clear that mutual forbearance requires some minimum level of concentration in the

markets involved to be a plausible motive. In fragmented markets, existing or potential

competition from firms lacking the "proper" combinations of market presence are likely to

force multimarket firms into competing vigorously.

5.2 Relatedness Measures and Mutual Forbearance

The possibility of creating and exploiting gains from mutual forbearance may be relevant for

both the entry and exit decisions of diversified firms.

Firms may refrain from exiting a weak position in one industry because maintaining this

position is important to protect gains from mutual forbearance in another. Or in other words;

the firm remains in the weak business, not because it is reaping gains from relatedness, but

because remaining there is a perquisite for low levels of rivalry elsewhere. As an illustration,

assume that firm A is competing in two different industries, 1 and 2. Assume further that the

two main competitors, B and C, are also present in both industries. Firm A is considering

leaving industry 2, but may refrain from doing so for fear of being attacked by B or C in

industry 1. If this is the outcome, the decision to stay in industry 2 is not a result of

relatedness, but of an attempt to protect gains from mutual forbearance in industry 1. A

researcher studying this situation using a survivor-based measure of relatedness would

register that industries 1 and 2 are frequently combined inside firms, indicating that they are

related. Firm A's decision to stay in industry 2 may therefore falsely be interpreted as

evidence of the benefits of combining the related industries 1 and 2. Using a SIC-based

measure reduces the risk of drawing such a false conclusion, since overlapping patterns of industry participation is not built into the SIC-based measures.

Based on this one may speculate that the findings reported in Lien (2003a) - that survivor-based measures are superior to SIC-based measures in terms of explaining exit decisions - is an artifact of survivor-based measures capturing the mutual forbearance motive better, and not evidence of a superior ability to measure relatedness. If this is the case, we would expect the survivor-based measures to be superior in terms of predicting exit decisions involving fairly concentrated industries (where the mutual forbearance motive is plausible), but not to exist for industries with low concentration. We can then formulate the following hypothesis:

H3:   Survivor-based measures of relatedness will outperform SIC-based measures of

relatedness in terms of predicting exit decisions when industry concentration is high,

but not when it is low

Now consider the case of entry decisions. Firms may prefer to enter industries where it will meet existing competitors as a mean of establishing mutual forbearance. Assume for example that firm A is currently only present in industry 1, but fears an attack by its two most prominent competitors B and C. Assume that B and C are both competing in industry 2. One possible avenue for discouraging an attack in industry 1 is for A to establish a foothold in industry 2 and threaten to retaliate there. Thus the entry by A into industry 2 is not driven by relatedness between industries 1 and 2, but to create a situation of mutual forbearance in industry 1. Again, a researcher studying this situation using a survivor-based measure of relatedness may conclude that industries 1 and 2 are related, and that firm A's entry decision is driven by a desire to exploit this relatedness. Using a SIC-based measure reduces the risk of

drawing such a false conclusion, since overlapping patterns of industry participation is not an element in the construction of the SIC-based measures.

Based on this one may speculate that the findings reported in Lien (2003b) - that survivor-based measures are superior to SIC-based measures in terms of explaining entry decisions - results from survivor-based measures capturing the mutual forbearance motive better, and not a superior ability to measure relatedness. If this is the case, we would expect the survivor-based measures to be superior in terms of predicting entry decisions involving fairly concentrated industries (where the mutual forbearance motive is plausible), but not to exist for industries with low concentration (where the mutual forbearance motive is not plausible). We can thus formulate the following hypothesis:

H4:   Survivor-based measures of relatedness will outperform SIC-based measures of relatedness in terms of predicting entry decisions when industry concentration is high, but not when it is low

6.   Testing the Mutual Forbearance Hypothesis

6.1 Data and Methodology

Samples

The data used to test H3 and H4 are the same samples as were used by Lien (2003a and 2003b), which is advantageous because it allows us to keep everything constant but the variable of interest - which as noted in H3 and H4 is the level of industry concentration.

The exit sample used to test H3 is described in greater detail in Lien (2003a), but we note here that this sample consists of 593 instances of exit and 598 instances of randomly chosen nonexits made between 1981 and 1985. The total sample thus includes a total of 1191

observations. These were obtained from 70 different parent firms that had sales exceeding $ 20 millions in 1981, that were not liquidated in their entirety between 1981 and 1985, that had data in both the Trinet and Compustat data files, and where undisputable matches between the two databases could be obtained.

The entry sample used to test H4 is described in greater detail in Lien (2003b). It consists of 1202 instances of entry and 1176 instances of non-entry (randomly chosen) occurring between 1981 and 1985. The total sample thus consists of 2378 observations, representing 145 parent firms. The criteria for selection were the same as for the exit sample with the exception of minimum sales being set to $ 10 millions (in 1981).

Statistical Methods

To test H3 and H4 we divided both the exit- and the entry samples into two equally sized subsamples, one containing the most concentrated industries and one containing the least concentrated industries. We then reran the logistic regression analyses performed by Lien (2003a and 2003b) to examine whether his findings hold for both the high- and low concentration subsamples. The general model for the logistic regression analyses were the following:

P(Exit=1) = β_1 + β_2(Industry growth) + β_3(Industry concentration) + β_4(Industry profitability) + β_5(Parent size) + β_6(Parent market share) + β_7(Parent leverage) +β_8(Parent liquidity) + β_9 (Parent relatedness) + ε

P(Entry=1) = β_1 + β_2(Industry growth) + β_3(Industry concentration) + β_4(Industry profitability) + β_5(Parent size) + β_6(Parent diversity) + β_7(Parent relatedness) + ε

Independent Variables

Testing whether the findings in Lien (2003a and 2003b) hold for both the high and low concentration subsamples dictates that the independent variables are held constant, so we use exactly the same variables as did he.

The test of H3 (the exit sample) involves four different independent variables, two survivor-based measures (SURVTOT and SURVCORE), and a SIC-based equivalent to each of these, SICTOT and SICCORE respectively. The first of theses pairs, SURVTOT and SICTOT have already been defined in section 2.2 (SURVTOT) and section 4.1 (SICTOT), see also Lien (2003a). As noted these measure the weighted average relatedness of an industry i to all other industries participated in by the parent, based on survivor- and SIC-measures respectively. The second of these pairs, SURVCORE and SICCORE, measure how related a business i is to the core business c of the parent, where the core business is defined as the largest 4-digit SIC code in terms of sales. The measure SURVCORE was also defined in section 2.2 above, which leaves only SICCORE undefined. SICCORE is defined in the following manner:

$$SICCORE_i = d_{ic} \frac{s_i}{\sum_i s_i}$$

Where $d_{ic} = 3$ if $i = c$

$d_{ic} = 2$ if $i \neq c$ and i and c are in the same 3-digit SIC codes

$d_{ic} = 1$ if i and j are in different 3-digit SIC-codes, but similar 2-digit SIC codes

$d_{ic} = 0$ if i and c are in different 2-digit SIC codes

The test of H4 (the entry sample) also involves four different independent variables. Two survivor-based measures (SURVTOT and SURVNBOR) and their respective SIC-based equivalents (SICTOT and SICNBOR). All of these have been defined in sections 2.2 and 4.1

of this paper, but we restate that the variables with the suffix -NBOR measure the weighted relatedness of a given business i to the two closest neighboring businesses of the parent (see also Lien 2003b).

Control Variables

To keep as much as possible constant when comparing the present findings to the original studies by Lien, we use exactly the same control variables as did he. In the interest of brevity, and since the behavior of the control variables is not the issue of interest here, we will not conduct a detailed discussion of control variables.[48] The control variables included are:

INDGRSAL. The growth in percent of industry sales between 1981 and 1985. Source: Trinet

INDCONC. Four firm concentration ratio for 1981. Source: Trinet

INDPROF. Median ROA for each industry over the period 1980-1982. Source:Compustat

MKTSH. Firm sales in industry i as percent of industry sales in 1981. Source: Trinet

PARSIZE. Total sales of parent in 1981 Source: Trinet

PARDIV. Number of four digit SIC-codes participated in by a parent in 1981. Source: Trinet

PARLEV. Long term debt to market value in 1981. Source: Compustat

PARLIQ. Ratio of current assets to current liabilities in 1981. Source: Compustat

Dependent Variables

The dependent variable involved in testing H3 is exit. If a parent active in a four-digit SIC code in 1981 has exited this industry by 1985, the dependent variable is given a value of 1. If the parent is still active in the industry, the value assigned is 0. Both divestures and closures are thus considered to represent exit.

---

[48] We refer to Lien (2003a and 2003b) for a more detailed discussion of the control variables

The dependent variable involved in testing H4 is entry. If a parent entered a four-digit SIC code it was not present in in 1981 by 1985, the dependent variable is given a value of 1. If the parent the parent did not enter the industry in question, the value assigned is 0.

6.2   Results and Discussion

H3 stated that the superior performance of the survivor-based measures in terms of predicting exit would only hold for the high concentration sample. Table 5 shows the logistic regression outputs for both the high and low concentration subsamples. The regression results strongly contradict H3. Not only does the superiority of the survivor-based measures hold for both the high and low concentration subsamples, it is actually stronger for the low concentration sample than for the high concentration sample. While this can be seen from all the reported measures of model performance, it is perhaps most clearly visible from the comparisons of changes in model chi-square as we move from model 2 to model 3 and from model 4 to model 5. These changes in model chi-square are not only positive and highly significant for both subsamples, but they are actually larger for the low concentration subsample than for the high concentration subsample. This is in stark contrast to H3, and leads us to conclude that the superior performance of the survivor-based measures of relatedness found by Lien (2003a) is not explained by contamination from mutual forbearance motives.

H4 made a similar prediction to H3 only it relates to data on entry - not exit. In other words it states that the superior performance of the survivor-based measures in terms of predicting entry would only hold for the high concentration sample. Table 6 shows the logistic regression outputs for both the high and low concentration subsamples. The regression results strongly contradict H4. The pattern is again that not only does the superiority of the survivor-based measures hold for both subsamples, it is even stronger for the low concentration sample than for the high concentration sample. This is also quite inconsistent with the suggestion that

the superior performance of the survivor-based measures is a result of unintended

contamination by mutual forbearance motives.

In sum, the strong contradiction of both H3 and H4 does seem to rule out the interpretation

that the superior performance of the relatedness measures found by Lien (2003a, 2003b) is

explained by mutual forbearance effects. It does not, however, indicate that mutual

forbearance does not exist, nor that survivor-based measures are uninfluenced by mutual

forbearance motives. What these findings do seem to suggest is that any such noise in the

survivor-based measures is probably small compared to their ability to better capture actual

relatedness.

7. Conclusions

For empirical research on corporate strategy to advance, a crucial challenge is to find better

ways of measuring the key theoretical constructs. Few if any constructs are more central

theoretically than relatedness, and few if any measurement procedures have been more

heavily criticized than those used to capture relatedness (Fan and Lang, 2000; Markides and

Williamson 1994, 1996; Robins and Wiersma 1995, 2003; Silverman, 1999). The survivor-

based approach does surely not represent the end of this discussion, but the results by Lien

(2003a, 2003b) along with those reported in the present paper does strongly suggest that the

survivor-based approach represents a step forward in comparison with the conventional SIC-

based measures.

The contribution of this paper has been to examine whether the superior ability of the

survivor-based measures to explain the entry and exit decisions of diversified firms could be

spurious, resulting from a correlation between survivor-based measures and other influences

on behavior than that of exploiting relatedness. En ante, the two most likely candidates to do

so are the herd behavior mechanism - and motives related to mutual forbearance. The reason being that the way the survivor-based measures are constructed makes them well suited to capture such effects. However, the data reported here shows no sign that such spuriousness is an important source of noise in the survivor-based measures, and they strongly contradict any suggestion that their apparent superiority over the SIC-based alternatives is entirely caused by this.

Having said this, we do not claim to given definite proof of the superiority of the survivor-based approach. It would still be desirable to complement the existing evidence with data from other periods, and to validate the survivor-based measures against other dependent variables than entry and exit. A logical next step could for example be to use performance data, such as revenue growth, return on assets, Tobins q or other measures of market value, and examine whether survivor-based measures explain more of the variation in these variables than rival measures of relatedness. The importance of the relatedness concept along with the encouraging evidence accumulated so far should make such additional efforts well worth their while.

REFRENCES

Alchian AA. 1950. Uncertainty, Evolution, and Economic Theory. *Journal of Political Economy* LVIII: 211-221.

Banerjee AV. 1992. A Simple-Model of Herd Behavior. *Quarterly Journal of Economics* 107(3): 797-817.

Barnett WP. 1993. Strategic deterrence among multipoint competitors. *Industrial and corporate Change* 2: 249-278.

Barnett WP, Greve HR, Park DY. 1994. An Evolutionary Model of Organizational Performance. *Strategic Management Journal* 15: 11-28

Baum JAC, Korn HJ. 1996. Competitive dynamics of interfirm rivalry. *Academy of Management Journal* 39: 255-291.

Baum JAC, Korn HJ. 1999. Dynamics of dyadic competitive interaction. *Academy of Management Journal* 20: 251-278.

Baum JAC. 1999. The rise of chain nursing homes in Ontario. *Social Forces* 78: 543-584.

Baumol WJ, Panzar JC, Willig RD. 1982. *Contestable markets and the theory of industry structure*. Harcourt Brace Jovanovich: New York.

Bulow JI, Geanakoplos JD, Klemperer PD. 1985. Multimarket Oligopoly - Strategic Substitutes and Complements. *Journal of Political Economy* 93(3): 488-511.

Edwards CD. 1955. Conglomerate bigness as a source of power, *NBER conference report Business Concentration and Economic Policy*. Princeton University Press.: Princeton.

Elster J. 1989. *Nuts and Bolts for the Social Sciences*. Cambridge University Press: New York.

Evans WN, Kessides IN. 1994. Living by the Golden-Rule - Multimarket Contact in the United-States Airline Industry. *Quarterly Journal of Economics* 109(2): 341-366.

Fan JPH, Lang LHP. 2000. The measurement of relatedness: An application to corporate diversification. *Journal of Business* 73(4): 629-660.

Feinberg RM. 1985. Sales-at-Risk - a Test of the Mutual Forbearance Theory of Conglomerate Behavior. *Journal of Business* 58(2): 225-241.

Friedman M. 1953. The Methodology of Positive Economics, *Essays in Positive Economics*: 328. University of Chicago Press: Chicago.

Fudenberg D, Tirole J. 1984. The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look. *American Economic Review* 74(2): 361-366

Gimeno J, Woo CY. 1996. Hypercompetition in a multimarket environment: The role of strategic similarity and multimarket contact in competitive de-escalation. *Organization Science* 7(3): 322-341.

Gimeno J. 1999. Reciprocal threats in multimarket rivalry: Staking out 'spheres of influence' in the US airline industry. *Strategic Management Journal* 20(2): 101-128.

Greve HR, Baum JAC. 2001. Introduction: A multiunit, multimarket world, *Advances in Strategic Management*, Vol. 18: 1-28.

Hodgson G. 1993. *Economics and Evolution*. University of Chicago Press: Ann Arbor.

Keynes JM. 1936. *The general theory of employment interest and money*. Macmillan: London.

Lien LB. 2003a). Yet another way of measuring relatedness - This one: Let competition do it! *This thesis paper no. 2.*

Lien LB. 2003b). Relatedness and patterns of diversification: A survivor-based approach. *This thesis paper no.3.*

Markides CC, Williamson PJ. 1994. Related Diversification, Core Competence and Corporate Performance. *Strategic Management Journal* 15: 149-165.

Markides CC, Williamson PJ. 1996. Corporate Diversification and Organizational Structure: A Resource-Based View. *Academy of Management Journal* 39: 340-367.

Phillips OR, Mason CF. 1996. Market regulation and multimarket rivalry. *Rand Journal of Economics* 27(3): 596-617.

Robins J, Wiersema MF. 1995. A Resource-Based Approach to the Multibusiness Firm - Empirical-Analysis of Portfolio Interrelationships and Corporate Financial Performance. *Strategic Management Journal* 16(4): 277-299.

Robins JA, Wiersema MF. 2003. The measurement of corporate portfolio strategy: Analysis of the content validity of related diversification indexes. *Strategic Management Journal* 24(1): 39-59.

Scharfstein DS, Stein JC. 1990. Herd Behavior and Investment. *American Economic Review* 80(3): 465-479

Scott JT. 1989. Purposive Diversification as a Motive for Merger. *International Journal of Industrial Organization* 7(1): 35-47.

Scott JT. 1991. Multimarket Contact among Diversified Oligopolists. *International Journal of Industrial Organization* 9(2): 225-238.

Silverman BS. 1999. Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. *Management Science* 45(8): 1109-1124.

Teece DJ, Rumelt R, Dosi G, Winter S. 1994. Understanding Corporate Coherence - Theory and Evidence. *Journal of Economic Behavior & Organization* 23(1): 1-30.

Winter SG. 1971. Satisficing, Selection, and Innovating Remnant. *Quarterly Journal of Economics* 85(2): 237-261.

Table 1: Variable definitions, data source and predicted sign

Variable	Definition	Data Source	Predicted Sign
SURVTOT	Weighted average *survivor-based* relatedness of industry *i* to all industries in the portfolio of the parent	Trinet	-
SURVNBOR	Weighted *survivor-based* relatedness of industry i to the two closest businesses of the parent	Trinet	-
SICTOT	Weighted average *SIC*- based relatedness of industry *i* to all industries in the portfolio of the parent	Trinet	-
SICNBOR	Wheighted *SIC*-based relatedness of industry i to the two closest businesses of the parent	Trinet	-

Table 2: Means standard deviations and correlation coefficients of independent variables (N=229)

	Variable	Mean	SD	Survtot	Survnbor	Sictot
1	Survtot	5,79	5,68			
2	Survnbor	17,60	10,54	0,54***		
3	Sictot	0,122	0,262	0,52***	0,22***	
4	Sicnbor	0,816	0,721	0,22***	0,33***	0,047***

*** Correlation is significant at the 0,01 level
** Correlation is significant at the 0,05 level
* Correlation is significant at the 0,1 level

Table 3: Logistic Regression Output with Different Relatedness Measures (N=229)

	SURVTOT	SURVNBOR	SICTOT	SICNBOR
Variable coefficient	-0,16***	-0,04***	-1,65**	-0,13
Wald statistic	16,64	6,80	4,31	0,43
Model Chi-square	23,47***	7,87***	5,90***	0,43
-2 log likelihood	266,17	281,77	283,74	289,21
Cox and Snell r²	0,097	0,034	0,025	0,002
Nagelkerke r²	0,136	0,047	0,035	0,003

Table 4: Logistic Regression Output: Models containing only SURVTOT, only SICTOT, and both. (N=229)

	SURVTOT only		SICTOT only		SURVTOT and SICTOT	
	Beta	Wald	Beta	Wald	Beta	Wald
SURVTOT	-0,16***	16,64			-0,162***	13,252
SICTOT			-1,653***	5,901	-0,030	0,01
Model Chi-square	23,47***		5,90***		23,47	
-2 log likelihood	266,17		283,74		266,17	
Cox and Snell r^2	0,097		0,025		0,097	
Nagelkerke r^2	0,136		0,035		0,136	

Table 5. Logistic regression output, exit sample   N =596 (subsample with high concentration)   N = 595 (subsample with low concentration)

Model	1. Controls only		2. Ctrs +survtot		3. Ctrs + sictot		4. Ctrs +survcore		5. Ctrs +sictot	
Subsample (concentration)	High	Low	High	Low	High	Low	High	Low	High	Low
	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)
Constant	0,51 (0,84)	0,60 (1,45)	1,13** (3,73)	0,97** (3,48)	0,50 (0,81)	0,60 (1,43)	0,79 (1,93)	0,89* (2,93)	0,63 (1,26)	0,67 (1,71)
Indgrsal	-0,86*** (29,99)	-0,29 (1,80)	-0,83*** (26,89)	-0,32 (1,99)	-0,85*** (29,08)	-0,30 (1,94)	-0,83*** (28,19)	-0,41** (3,16)	-0,84*** (28,84)	-0,35* (2,59)
Indconc	0,00 (0,37)	0,00 (0,01)	0,00 (0,32)	0,01 (0,42)	0,00 (0,39)	0,00 (0,06)	0,00 (0,17)	-0,00 (0,03)	0,00 (0,30)	0,00 (0,07)
Indprof	-0,08 (0,12)	0,00 (0,00)	-0,01 (0,00)	0,03 (0,01)	-0,07 (0,09)	0,02 (0,01)	-0,15 (0,38)	-0,12 (0,26)	-0,15 (0,39)	-0,06 (0,06)
Mktsh	-0,13*** (29,42)	-0,36*** (14,30)	-0,10*** (20,15)	-0,23*** (6,87)	-0,13*** (26,80)	-0,33*** (12,31)	-0,08*** (10,54)	-0,12 (1,87)	-0,10*** (17,88)	-0,24*** (6,23)
Parsize	0,00 (0,08)	0,00 (0,29)	0,00 (0,32)	0,00 (0,58)	0,00 (0,10)	0,00 (0,39)	0,00 (0,82)	0,00 (1,34)	0,00 (0,41)	0,00 (0,83)
Parlev	0,35 (0,11)	-1,39* (2,19)	0,70 (0,42)	-1,21 (1,58)	0,39 (0,14)	-1,30 (1,92)	0,34 (0,10)	-1,31 (1,81)	0,37 (0,13)	-1,19 (1,53)
Parliq	-0,04 (0,03)	0,02 (0,01)	-0,04 (0,03)	0,10 (0,27)	-0,03 (0,02)	0,04 (0,04)	-0,13 (0,34)	-0,02 (0,01)	-0,09 (0,18)	-0,02 (0,01)
SURVTOT			-0,08*** (21,43)	-0,09*** (28,49)						
SICTOT					-0,18 (0,21)	-0,65* (2,41)				
SURVCORE							-0,66** (4,37)	-1,86*** (11,91)		
SICCORE									-3,26* (2,59)	-55,04** (4,49)
-2Log likelihood	720,85	796,84	695,56	761,40	720,64	794,32	702,41	760,21	711,93	776,53
ModelChi-square	105,14***	27,92***	130,43***	63,36***	105,35***	30,44***	123,58***	64,55***	114,06***	48,24***
Cox and Snell r^2	0,162	0,046	0,197	0,101	0,162	0,050	0,187	0,103	0,172	0,078
Nagelkerke r^2	0,216	0,061	0,262	0,135	0,216	0,067	0,250	0,137	0,232	0,104
Δ Chi-square vs. model 1 (controls only)			25,28***	35,44***	0,21	2,52*	18,44***	36,63***	8,92***	20,32***

Δ Chi-square model 2 vs. model 3   High concentration sample: 25,07***   Low concentration sample: 32,92***

Δ Chi-square model 4 vs. model 5   High concentration sample: 9,52***   Low concentration sample: 16,32***

Table 6. Logistic regression output, entry sample   N =1189 (subsample with high concentration) N=1189 (subsample with low concentration)

Model	1. Controls only		2. Ctrs +survot		3. Ctrs + sictot		4. Ctrs +survnbor		5. Ctrs +sicnbor	
Subsample	High	Low	High	Low	High	Low	High	Low	High	Low
	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)
Constant	0,81*** (16,93)	0,21 (1,32)	-0,50** (4,40)	-0,83*** (14,60)	0,69*** (11,13)	-0,03 (0,03)	-0,86*** (12,43)	-0,99*** (19,83)	0,45** (4,42)	-0,13 (0,41)
Indgrsal	0,04 (1,24)	0,12** (4,87)	0,04 (1,19)	0,12** (4,56)	0,04 (1,37)	0,13*** (6,33)	0,04 (1,05)	0,06 (1,07)	0,04 (1,31)	0,09* (2,62)
Indconc	-0,02*** (34,03)	-0,01 (2,01)	-0,02*** (17,18)	-0,03*** (6,63)	-0,02*** (40,50)	-0,02* (2,80)	-0,01*** (11,83)	-0,03*** (7,41)	-0,02*** (31,05)	-0,01 (1,70)
Indprof	-0,06 (0,20)	0,08 (0,36)	-0,20 (1,99)	-0,02 (0,02)	-0,21* (2,63)	0,03 (0,05)	-0,07 (0,21)	-0,06 (0,16)	-0,19* (2,09)	0,04 (0,09)
Parsize	0,00 (0,01)	0,00 (0,01)	0,00 (0,93)	0,00** (4,78)	0,00 (0,45)	0,00 (0,15)	0,00 (0,24)	0,00 (0,43)	0,00 (0,12)	0,00 (0,18)
Pardiv	0,00 (0,01)	0,00 (1,03)	0,00 (0,09)	0,00 (0,27)	0,00 (0,00)	0,00 (1,46)	-0,01*** (12,42)	-0,01*** (8,48)	-0,01*** (10,66)	-0,01*** (6,32)
SURVTOT			0,37*** (193,90)	0,43*** (209,16)						
SICTOT					4,59*** (56,35)	4,35*** (45,92)				
SURVNBOR							0,15*** (206,3)	0,17*** (217,51)		
SICNBOR									1,37*** (178,64)	1,24*** (147,05)
-2Log likelihood	1606,23	1630,52	1278,96	1265,20	1504,13	1551,13	1304,40	1272,79	1385,37	1454,63
ModelChi-square	36,48***	7,78	363,54***	373,10***	138,34***	87,17***	338,11	365,51***	257,14***	183,66***
Cox and Snell r^2	0,03	0,007	0,263	0,269	0,110	0,071	0,248	0,265	0,194	0,143
Nagelkerke r^2	0,04	0,009	0,352	0,360	0,147	0,095	0,331	0,354	0,260	0,191
Δ Chi-square vs. controls only			327,06***	365,31***	101,89***	79,38***	301,62***	357,73***	220,65***	175,88***
Δ Chi-square model 2 vs. model 3					High concentration sample: 225,20*** Low concentration sample: 285,93***					
Δ Chi-square model 4 vs. model 5									High concentration sample: 80,97*** Low concentration sample: 181,85***	

CLOSING REMARKS

The very first words in this thesis (the foreword) stated that learning was the desired output. At this stage it seems natural to take stock, not of my own learning, but of what if anything the thesis has added to the field(s) of inquiry addressed.

Unfortunately, under a Popperian view of science we cannot ultimately say we know anything in general from empirical evidence. However, empirical evidence can make us less unsure about claims of knowledge. So what are we less unsure about after this work has been completed?

It is my opinion that we can be less unsure about the validity of relying on the SP as an empirical strategy in research on corporate diversification. We can also be less unsure about the validity of the existing research in corporate strategy that has employed the SP. We cannot, however, generalize this to say that relying on the SP is OK in any setting or for studying any issue. On the other hand the fact that the SP did survive testing in the context of corporate diversification should somewhat increase the a priori probability that it is valid in related areas - such as for example the study of vertical integration (where it is also commonly applied). On the other hand it is pertinent to repeat the call for more testing of the robustness of the SP. For example, one may argue that even though it seems valid in a large economy such as the US, it may not be equally valid in, say, Norway. And one may further argue that even though the findings supplied here indicate that it is valid for the private sector as a whole, there may be industries or settings within the private sector where it does not hold. Put differently, we have done more in terms of validating the SP for use on large inter-industry samples than for smaller intra-industry samples.

The second question raised was whether the SP could be useful for overcoming the problem of measuring relatedness. If we consider the starting point as one of complete ignorance as to whether the answer to this question is yes or no, we have gradually become less uncertain in the direction of an affirmative answer. Fortunately (or unfortunately) the benchmark in terms of the quality of existing measures is not of the toughest. Basically, what the three papers on this issue have done is putting obstacles in the way of those holding the opposite opinion. By demonstrating a superior ability to explain entry and exit decisions, antagonists have to find counterarguments consistent with these findings. And controlling for herd behavior and mutual forbearance interpretations, we have made the task of finding alternative interpretations that fit the data even more difficult. But difficult is not the same as impossible. There are still explanations that could fit the data, and not lead to the conclusion that the survivor-based approach is superior in terms of capturing relatedness. One candidate could be institutional theory (Dimaggio and Powell, 1983), where social and cultural pressures in combination with a need for legitimacy may cause behavior consistent with our data. And surely there may be other explanations that fit the data.

As always, the solution seems to lie in more empirical testing, and as repeatedly stated, the next logical step is to evaluate the survivor-based measures with respect to other measures of performance (such as ROA, tobins q, growth, etc.).

Are these findings important? Given that the hallmark of the scientific method is the idea of testing claims of knowledge against data, it is crucial that the data analyzed can speak to the truth of the hypotheses tested. If the SP was invalid, it would mean that a substantial amount of the data we have collected in industrial organization, organizational economics and strategic management was of questionable relevance for the hypotheses (about efficiency)

they were meant to test. Hence, a contribution that seeks to reduce the uncertainty about the validity of the SP is not trivial. The test supplied here is the first explicit attempt to do so. The search for better ways of solving the problem of measuring relatedness is important for the same reason. Claims of knowledge about the effects of relatedness on various other variables hang in the open as long as we cannot agree on whether relatedness has been captured by the measurement procedures used. Given the prominence of the relatedness variable, this problem is an important one for advancing knowledge about corporate strategy. Contributions towards solving this problem should therefore be welcome in the field. The survivor-based approach analyzed here is dramatically different from all other measures, and it has not been evaluated previously. This in combination with the promising findings contained in this thesis should be of some interest for future empirical research. Finally, we note that the survivor-based approach may be used to overcome other measurement problems in other areas of research. Stigler, for example, has used the SP in studying economies of scale (Stigler, 1968). There are probably other undiscovered uses of the SP as well. This, however, shall be left for others to explore. This thesis ends here.

REFERENCES

Dimaggio PJ, Powell WW. 1983. The Iron Cage Revisited - Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review* **48**(2): 147-160.

Stigler GJ. 1968. *The organization of industry*. R. D. Irwin: Homewood, Ill.