WHAT'S THE MATTER WITH RELATEDNESS?

Lasse B. Lien*
Norwegian School of Economics and
Business Administration
Breiviksveien 40
N-5045 Bergen
Norway
Tel: +47 55 959 726
Fax: +47 55 959 780
e-mail: lasse.lien@nhh.no

Peter G. Klein
University of Missouri
135 Mumford Hall
Columbia, MO 65211
USA
Tel: +1 573 882 7008
Fax: +1 573-882-3958
e-mail: pklein@missouri.edu

Abstract
A core proposition of the resource-based view of the firm is that related diversification is more efficient than unrelated diversification. Nevertheless, the empirical evidence is usually described as mixed or unstable. We empirically examine three possible explanations for the nature of these findings. One is that they reflect measurement problems. Another is that they reflect a failure to take the resource-based view far enough in terms of recognizing firm heterogeneity. The third is that the resource-based view is misspecified: relatedness is not a significant determinant of efficiency. We use a detailed line-of-business sample to examine these three explanations. Our findings clearly indicate that the main problem is measurement, and we describe an approach to overcoming these problems.

*Corresponding author

INTRODUCTION

The analysis of corporate diversification in strategic management is arguably dominated by the resource based view (RBV). One may further argue that the core prediction of the RBV is that related diversification should be superior to unrelated diversification, ceteris paribus. The logic is compelling and simple; when firms diversify into unrelated industries, they are less likely to reap benefits from resources in their other industries. At the extreme, a diversified entrant may face no advantages over an upstart firm (but larger costs due to added complexity). Theoretically, these ideas date back to at least Penrose (1959), and empirically at least to Rumelt's (1974) landmark study. Nevertheless, the current status is that most reviewers summarize the empirical findings regarding the benefits of related diversification as mixed or confusing (e.g. Hoskisson and Hitt, 1990; Lemelin, 1982; Markides and Williamson, 1994, 1996; Reed and Luffman, 1986; and Robins and Wiersema, 1995, 2003).[1]

This paper evaluates three possible explanations for why this core prediction of the RBV has not succeeded in producing consistent support. The first potential problem deals with measurement: relatedness is notoriously difficult to capture empirically, particularly in large-sample research. Relatedness is hard to measure because theory advices us that related diversification creates efficiency gains (over single-business firms and unrelated diversifiers) only if specific conditions are met, and these conditions are largely unobservable. Under this view, then, the mixed empirical support for the relatedness hypothesis is driven by the inability to capture, empirically, the conditions under which relatedness adds value. We call this the *measurement interpretation*.

A second possible explanation is that the empirical work does not take the RBV logic far enough. While the RBV-literature on competitive advantage stresses idiosyncratic, firm-specific

---

[1] Arend (2006) argues that the RBV, more generally, has not been subject to adequate empirical scrutiny, due to problems defining the key concept of resources, measuring the conditions under which resources generate rents, and specifying which reources are primary.

resources (Barney, 1991; Dierickx and Cool, 1989; Peteraf, 1993), the RBV literature on corporate diversification - particularly the empirical work - tends to focus on industry-level relatedness (Markides and Williamson, 1994, 1996). This involves downplaying intra-industry differences in resource bundles. One may therefore suspect that the kind of relatedness that matters for performance is idiosyncratic, firm-specific relatedness, and that this kind of relatedness is not closely correlated with industry-level relatedness. In other words, because firms differ in the resources they possess and managers differ in how they conceptualize relatedness (Stimpert and Duhaime, 1997), the benefits from combining a given pair of industries are highly idiosyncratic and firm-specific. Accordingly, one should not expect anything but mixed and inconsistent results when relatedness is treated as an abstract relationship between industries, unconnected to the specific characteristics of a given firm. Therefore, while we may gain little insight from focusing on the relationship between industries X and Y, it *is* meaningful to ask how the activities in industries X and Y are related in firm *Z* (Foss and Christensen, 2001). Under this interpretation, the mixed empirical findings are an artifact of the tendency to ignore firm-specific, unobservable characteristics. We call this the *idiosyncrasy interpretation*.

A third explanation is that the RBV is misspecified: relatedness, even correctly measured, is not a significant determinant of efficiency. Relatedness may not produce efficiency gains because managers are not pursuing efficiency, or because they systematically lack the information, cognitive ability, or organizational tools to realize efficiency gains. Expansion into related industries can for example be motivated by market power (via mutual forbearance), entrenchment of incumbent management, or the pursuit of other private benefits. Alternatively, the resource based view may assume too much rationality on the part of decision-makers to be empirically successful. Decision-makers may be prone to herding or simple errors, obscuring the link between relat-

edness and efficiency. We collectively label such non-efficiency explanations the *skeptic interpretation*.

We examine these alternative explanations using a comprehensive sample of line-of-business data from the 1980s. First we compare the measurement interpretation to the idiosyncrasy interpretation. Our first test asks whether decision makers act as if some combinations are inherently more beneficial than others and, if so, whether these combinations are effectively captured by conventional measures of relatedness. We find that some combinations are clearly and systematically perceived to be more attractive than others. In other words, the revealed preferences of decision makers indicate that relatedness is not predominantly firm specific. We also find that conventional approaches to measuring relatedness using to distances in the SIC system do not effectively capture the kinds of relatedness that decision makers care about.

A possible objection is that while the choice of combinations may be systematic, the performance resulting from these combinations is idiosyncratic. Our second test investigates the performance consequences of particular combinations. Is there a systematic relationship between relatedness and performance, or is performance firm specific? Moreover, if performance is systematically associated with relatedness, do conventional SIC-based measures capture this association effectively? Our results suggest that relatedness and performance are systematically associated, indicating that the performance effects of particular industry combinations are not entirely firm specific. Again, we also find that the SIC-based measures seem to do a poor job of capturing these associations. Taken as a whole, these results favor the measurement interpretation over the idiosyncrasy interpretation.

Next we pit the measurement interpretation against the skeptic interpretation. Even if relatedness is unimportant, could certain combinations of industries appear more frequently in the data due to herding (Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992) or mutual forbear-

ance (Edwards, 1955; Karnani & Wernerfelt, 1985)? The performance effects described above suggest that herding is not driving our empirical findings. To see if mutual forbearance explains the decision to diversity into certain industries but not others, we examine the relationship between relatedness and target-industry concentration. Mutual forbearance provides a motive for entry only if industry concentration exceeds some minimum threshold. If our results are driven by a mutual forbearance motive, relatedness should be a better predictor of entry into concentrated industries than more fragmented ones. We find that this is not the case.

Overall, our evidence suggests that there are significant and stable efficiency effects of relatedness, and that decision makers do act in a manner consistent with the pursuit of such effects. Contrary to the idiosyncrasy interpretation, these efficiency effects are not entirely firm specific, so that looking at inter-industry relatedness makes sense. Moreover, traditional approaches to relatedness using SIC distances seem to do a poor job of capturing the relatedness that drives diversification decisions and their outcomes. While both the idiosyncrasy and skeptic interpretations have merit, our results point to measurement problems as the main reason why empirical research on relatedness has produced mixed and confusing findings.

THE THEORETICAL CASE FOR A MEASUREMENT INTERPRETATION

We suggested above that relatedness matters only when particular theoretical conditions are met, and that conventional measures of relatedness may not capture these conditions very well. The theory of scope economies suggests that three basic conditions must be met for relatedness to create value. First, the diversifying firm must possess some resources that are functional substitutes for the resources in the target industry. Second, these resources must be at least partly indivisible, leaving single-business firms and unrelated diversifiers with costly excess capacity (Penrose, 1959). If these resources are not fully exploited in the firm's existing businesses, they may

be deployable at low marginal cost in a new business. Conversely, without excess capacity, a diversified firm would have no advantages over an independent startup and could face entry costs not born by incumbents in the destination industry.

Third, there must be transaction costs in the market for excess capacity. As Teece (1980, 1982) points out, while the existence of such 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 markets, single-business firms and unrelated diversifiers can simply sell or rent out their excess capacity, or buy the capacity they need from other firms. In other words, absent transactional difficulties, two separate firms could simply contract to share the inputs, facilities, or whatever accounts for the relevant scope economies. If they do not, it must be because the costs of writing or enforcing such a contract are greater than the benefits from joint production. Whether the firms will integrate thus depends on the comparative costs and benefits of contracting, not on the underlying production technology. Indeed, if contracting costs are low, the related diversifier may actually compete at a disadvantage relative to the single-business firm, because the diversified firm faces the additional bureaucratic costs of low-powered incentives, increased complexity, and so on (Williamson, 1985).[2]

More recent attention has focused not on resource substitutability, but resource complementarity (Christensen and Foss, 1997; Foss and Christensen, 2001; Larsson and Finkelstein,

---

[2] Recent literature emphasizes *dynamic* complementarities, the ability to identify new ways of combining existing resources or speed up the development of new resources. The benefits to similarity in this context arise because such dynamic complementarities may be greater if the industries in question share some basic features (March, 1991), or because some common characteristics facilitate their exploitation (Finkelstein and Haleblian, 2002; Prahalad and Bettis, 1985). The degree of dynamic complementarity between industries thus depends on the balance between variety and similarity (Christensen and Foss, 1997). Industries with appropriate balances between variety and similarity produce larger dynamic complementarities than industries that are too different or too similar. Empirically, this implies that portfolios of businesses with strong inter-industry complementarities should be considered related (or, in Teece et al.'s [1994] language, "coherent"), and that firms with related activities should outperform firms with unrelated combinations of activities and single-business firms, ceteris paribus (again, assuming positive contracting costs).

1999; Teece, Rumelt, Dosi and Winter, 1994). Complementarities exist when investment in one industry increases the value of resources used in another industry, or when decisions about resource use in one industry affect similar decisions in another. These positive spillovers create a quantitative and qualitative coordination problem which may be best managed within a diversified firm (Milgrom and Roberts, 1992; Richardson, 1972). To explain why this coordination problem cannot be solved in the market (i.e., between single-business firms or unrelated diversifiers), we must appeal to some form of contracting costs. Hence, transaction costs are also relevant to situations involving complements.

In sum, similarity among a diversified firm's industries should not provide efficiency advantages unless accompanied by indivisibilities (either in the form of productive capacity or positive spillovers) and transaction costs. Unfortunately, as discussed below, capturing these conditions empirically is difficult.

Problems with conventional measures of relatedness

The most common measures of relatedness in existing research are categorical measures and continuous SIC-based measures. However, neither effectively captures the conditions described above under which relatedness should lead to performance improvements.

Categorical measures. The categorical approach, most closely associated with Rumelt (1974), is perhaps the best-known attempt to capture relatedness. Based on three ratios, Rumelt classified diversification strategies into four broad categories (nine if subcategories are included): single-business firms, dominant-business firms, related firms, and unrelated firms. The ratios used for classification are the specialization ratio (the proportion of a firm's revenue attributable to its largest single business), the related ratio (the proportion of a firm's revenue attributable to its largest group of related businesses), and the vertical ratio (the proportion of a firm's revenue aris-

ing from all byproducts, intermediate products, and end products of a vertically integrated sequence of processing activities).

A drawback of this procedure is that the classification of businesses into related and unrelated—needed to compute the related ratio—is done subjectively, using similarities in inputs, production technology, distribution channels and customers (Markides and Williamson, 1996; Robins and Wiersema, 1995). Besides the reliability of the subjective element in the classifications, this procedure also measures relatedness on a nominal level, only allowing comparisons within group averages.[3]

Moreover, this procedure mainly captures the degree to which resources are potential substitutes across industry boundaries, which is only the case when there are large similarities in inputs, production technology, distribution channels and customers. These measures do not capture indivisibilities, which determine whether excess capacity is likely to develop, nor do they incorporate any notion of transaction costs impeding coordination between firms. Because both indivisibilities and transaction costs are necessary for economies of scope to benefit related diversifiers, these measures are 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).

Continous SIC-based measures. Continuous SIC-based measures are currently the most popular relatedness variables in the strategy literature (Robins and Wiersema, 2003). These include the entropy index (Jaquemin and Berry, 1979), the concentric index (Caves, Porter and Spence,

---

[3] Determining the extent to which activities are vertically related also involves subjective judgment by the researcher.

1980), and similar metrics. Compared to Rumelt's categorical measures, these are less reliant on subjective judgments about the relatedness of particular activities are related. The assignment of SIC codes to activities by the Census Bureau does of course involve some subjectivity, of course, but the subjective component is consistent across studies that use SIC codes to compute relatedness. SIC-based measures also allow relatedness to be measured in intervals. 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.

However, the use of distances in the SIC system also introduces problems. It assumes that industries are homogenous within category levels, which is problematic if the breadth of the industry classifications vary, as is likely (Robins and Wiersema, 1995; Rumelt 1982). It also assumes that industries equally distant within the SIC hierarchy are equally dissimilar.

Most important for present purposes, however, SIC-based measurements are no better than categorical measures in capturing of indivisibilities and transaction costs.[4] In other words, even if distance among SIC codes is a good proxy for resource substitutability, these measures will tend to exaggerate relatedness. Foss and Christensen (2001) also point out that SIC-based procedures have an implicit bias towards economies of scope and thus are unlikely to capture dynamic complementarities, suggesting that this type of relatedness is prone to be underestimated. For these reasons, continuous SIC-based measures may be ineffective in capturing the relevant kinds of relatedness.

In sum, both categorical and continuous distance-based relatedness measures are problematic. The measurement explanation suggests that this is the reason for the mixed and often contradic-

---

[4] A study that does consider excess capacity is Chatterjee and Wernerfelt (1991).

tory findings in the empirical literature on relatedness. Below we propose an alternative measure, survivor-based relatedness, that mitigates these concerns.

A SURVIVOR-BASED APPROACH TO RELATEDNESS

Before turning the empirical tests we introduce an alternative measure of relatedness. The ideal measure features three characteristics. First, it should allow us to distinguish firm-specific (idiosyncratic) relatedness from relatedness that is a general property of the relationship between a given pair of industries. Second, it should form a useful benchmark for evaluating the potential problems with the conventional measures. Third, it should allow us to distinguish between efficiency and non-efficiency explanations. Here we present a survivor-based measure (SBM) of relatedness that we think can perform all these tasks.

The SBM is based on the survivor principle (SP). The core of the SP is 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)).[5] Two key processes ensure. One is that firms making negative profits will, unless some corrective measure is taken, lose resources and ultimately become extinct, while firms making positive profits will acquire resources and grow. The second assumption is that the desire for profit 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,[6] research in industrial organization, organizational economics, and strategic management suggest that it can be highly effective. After all, theories or hypotheses about what is efficient are rou-

[5] For the view that the competitive process creates outcomes that are optimizing, see Friedman (1953).

[6] Friedman (1953) is a possible exception.

tinely 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 the hypothesis that

vertical integration is more efficient than market governance when asset specificity is high is

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

(Shelanski and Klein, 1995; David and Hahn 2004; Klein, 2005).[7] 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; Zenger

and Marshall, 2000). 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 survivor principle holds, relatedness can be measured by observing what industries are

most often combined by firms in competitive markets. In other words, related activities are de-

fined as those most often performed together. Specifically, we estimate how much the frequen-

cies of actual combinations of 4-digit SIC industries deviate from what one would expect if di-

versification patterns were random. We take this difference to constitute a survivor-based meas-

ure of the relatedness between a pair of industries.

The SBM offers several potential advantages. First, it incorporates the knowledge of those

making portfolio decisions, presumably those with the best knowledge of the relevant benefits

and costs of combinations. Even if this knowledge is imperfect, or firms are rife with agency

---

[7] Some newer papers in the empirical transaction cost literature do not assume the SP, but rather employ a two-stage procedure in which in which the relationship between transactional characteristics and governance structure is endogenously chosen in the first stage, then used to explain performance in the second stage. Silverman, Nickerson and Freeman (1997), for example, show that transaction cost efficiency is positively correlated with firm survival in the for hire trucking industry, while Bigelow (2001) examines outsourcing arrangements in the U.S. automobile industry and finds that transactions that are appropriately aligned tend to last longer than inappropriately organized ones.

problems, these decisions have been screened by the competitive process, which tends to reverse poor decisions. Second, among the feasible alternatives, the survivor-based approach is holistic and flexible. It is holistic in the sense that it potentially captures all aspects of relatedness that are important for competitive outcomes, and flexible in the sense that it allows the causes of related-ness to vary across situations. As such, while the survivor-based approach makes no attempt to pinpoint what relatedness *is,* it is potentially useful for studying what relatedness *does*—i.e., how relatedness influences other variables of interest (performance, entry mode, financing decisions, organizational parameters, etc.).[8]

For our purposes the survivor-based approach has additional advantages. First, it potentially allows us to distinguish between firm-specific relatedness and general, inter-industry relatedness. This is because the SBM for a given firm is constructed from information about surviving combi-nations (of industries) in other firms, in a period immediately preceding the study period. If relat-edness is firm specific, the experiences of other firms should not reveal useful information about the benefits of similar combinations for the firm in question. This provides an opportunity to evaluate the idiosyncrasy interpretation. Second, the SBM can shed light on the size of measure-ment problems. If substituting the SBM for the conventional SIC-based measures of relatedness increases our ability to explain the variables that relatedness supposedly affects, this suggests that SIC-based measures are not capturing these effects effectively. This allows us to test the meas-urement interpretation. Finally, we can test the skeptic interpretation by examining if combina-tions are driven by factors other than efficiency, even when using the SBM to capture relatedness.

---

[8] Survivor-based measures could potentially help explain what relatedness is, however, by using measures of survi-vor-based relatedness as the dependent variable, and examining hypotheses about causes of relatedness (i.e. variables that explain relatedness).

Having presented the potential advantages of this approach, we do acknowledge that decision-makers make mistakes, or fail to act in firm owners' interests, and the competitive selection process is not perfect. For this reason, a survivor-based measure will include noise. Critics of the SP believe this noise is substantial (Elster, 1989; Hodgson, 1993; Winter 1971). All measures of relatedness are noisy, however. The more interesting question is whether a SBM is more or less noisy than feasible alternatives. The best judge of this is data.

Our approach here is based on a procedure originally developed by Teece, et al. (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:

$$\Pr(X_{ij} = x) = f_{hg}(x, K, n_i, n_j) = \frac{\binom{n_i}{x}\binom{K-n_i}{n_j-x}}{\binom{K}{n_j}} \tag{1}$$

The mean and variance of X_{ij} are:

$$\mu_{ij} = E(X_{ij}) = \frac{n_i n_j}{K}, \tag{2}$$

$$\sigma^2 = \mu_{ij}\left(1 - \frac{n_i}{K}\right)\left(\frac{K}{K-1}\right). \tag{3}$$

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:

$$SR_{ij} = \frac{J_{ij} - \mu_{ij}}{\sigma_{ij}} \tag{4}$$

The measure SR_{ij} is thus a standardized measure of how much the actual number of combinations exceeds 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. In this paper we focus on a measure that captures the weighted average relatedness of the focal industry i to all other businesses in the parent portfolio.[9] Assume a diversified firm that participates in m industries. Its business in industry j has sales of s_j and survivor-

---

[9] We have experimented with several other survivor based measures. One alternative does not consider how related the focal industry i is to *all* other business in the corporate portfolio, but how related it is to the *two closest* neighboring businesses of the parent. Yet, another focuses on the relationship between the focal business and the parents core business. Using these alternative measures reproduces substantively similar results (Reference withheld).

based relatedness SR_{ij} with industry i. The weighted average survivor-based relatedness ($SURVTOT_i$) of the target industry i to all other businesses in the firm is then defined as:

$$SURVTOT_i = \frac{\sum SR_{ij} s_j}{\sum s_j} \qquad (5)$$

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

$$SICTOT_i = \frac{\sum d_{ij} s_j}{\sum s_j} \qquad (6)$$

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

TEST 1: REVEALED PREFERENCES

Hypotheses

The first hypothesis uses decision-makers' revealed preferences to see whether particular combinations are systematically chosen, or whether relatedness is treated by market participants as idiosyncratic and firm specific. Since the measure $SURVTOT_i$ is built on a procedure that extracts information about relatedness from other firms' actions, this measure should not incorporate the firm-specific components of relatedness. Put differently, if the idiosyncracy interpretation is correct, $SURVTOT_i$ should be a poor predictor of entry. Thus we have:

H1: $SURVTOT_i$ is a poor predictor of the entry decisions of diversified firms.

The second hypothesis examines the measurement interpretation. If $SURVTOT_i$ is a useful predictor of entry decisions (i.e., H1 is not supported), we can ask what the revealed preferences of decision makers indicate about the usefulness of the conventional SIC-based measures. If $SURVTOT_i$ explains entry decisions significantly better than $SICTOT_i$, then the forms of relatedness that decision makers acting upon is poorly captured by SIC-based approaches. If the measurement interpretation is correct we should have:

H2: $SURVTOT_i$ explains the probability of entry significantly better than $SICTOT_i$

Methods Test 1

Testing H1 and H2 involves two distinct empirical operations. First we calculate the survivor-based measure of relatedness SR_{ij} for all possible pairs of industries in the US economy. Using this measure we calculate the variable $SURVTOT_i$ for any specific business belonging to any specific parent, and for any potential destination industry for any given parent. The second empirical operation is to test our hypotheses linking $SURVTOT_i$ and $SICTOT_i$ to the probability of a given parent entering a given industry.

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, including variables such as 4-digit SIC code, corporate ownership, and sales.[10] By ag-

---

[10] The Trinet database contains some sales figures that are imputed from multiplying employee counts with average industry sales per employee. To examine whether this constitutes a substantial source of error for our sample, we correlated the sales data from Trinet with Compustat data. This resulted in a correlation of 0,893 which indicates that

gregating 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 4-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 (Voigt,

1993), and that omissions are most likely for small firms (which are less likely to be diversified).

The primary measure of SR_{ij} is 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 results 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 are observed. The measure of SR_{ij} between the observed industry

pairs ranges from −7.97 to 93.55 with a mean of 4.33 and a standard deviation of 5.06. We also

calculate SR_{ij} for industry pairs that were not combined by 1981, because some of these where

combined in the subsequent periods where we observed entry or non-entry, and because some of

the randomly chosen non-entries created a need to calculate SR_{ij} for industry pairs that were never

combined. (For unobserved combinations, we simply set the number of observed combinations

(J_{ij}) to zero in the formula for calculating SR_{ij}.) Based on these calculations of SR_{ij}, we calculate

measures of $SURVTOT_i$ and $SURVNBOR_i$ by following the procedures described in the previous

section.

Sample for testing Hypotheses 1 and 2. The sample for testing H1 and H2 is derived as fol-

lows. We start with all 13,164 diversified firms in Trinet. To obtain the necessary data for the

variables of interest, Trinet data are merged with financial data from Compustat database. Since

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

---

the sales data in Trinet are of acceptable quality. The problem of imputed sales data is likely to be largest for the
smallest firms in the Trinet database (which are omitted from our sample).

phanumerically by parent name. Spelling differences between the two databases result in undisputable matches for 854 companies that had entries in all the years needed to compute the variables. To check that our sample is an unbiased sample of Compustat firms, we compare the average sales of our sample of matched firms to the average sales of the relevant population of diversified firms in the Trinet database, and find that the difference in sample means is not statistically significant.[11]

We imposed three further restrictions on our sample. First, we remove all firms with sales below $10 million. Second, we remove firms that where sold or liquidated in their entirety between 1981 and 1985. Third, we remove firms that did not enter at least one new 4-digit SIC code between 1981 and 1985. This reduces the sample to 145 firms. These 145 firms operated 4,582 businesses in 724 different 4-digit SIC codes in 1981. They entered 1,202 industries between 1981 and 1985, exited 1,118 industries, and remained in 3,464 industries throughout the period.

To test our hypotheses we included all the 1,202 instances of entry, but rather than using the entire sample of non-entries, we randomly select a matched sample of non-entries. 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 & Manski, 1981). Adjusting for 26 cases for which data were missing, this resulted in a final sample of a total of 2,378 observations, 1,202 entries and 1,176 non-entries.

Statistical methods. To test Hypotheses 1 and 2 we develop a model of the relationship between the probability of entry and relatedness, which controls for other potential motives for en-

---

[11] The mean sales of the firms in our matched sample is 63,713 with a standard deviation of 62,907. The mean of the relevant population in the Trinet database is 51,213, with a standard deviation of 51,587. The slightly larger size of the firms in our sample is probably due to the fact Compustat only includes publicly traded firms, while Trinet contains public and private firms. Publicly traded firms are on average larger than privately held firms, which means that using Compustat data will tend increase average firm size compared to using Trinet data alone.

try such as exploiting market power or joining a high-growth or high-return industry, and to re-
duce the risk that the performance of one or more of the presented relatedness measures are in-
flated or deflated because of associations with "other" variables. The general model is the follow-
ing:

P (Entry=1) $= \beta_1 + \beta_2$ (industry growth) $+ \beta_3$ (industry concentration) $+ \beta_4$ (industry profitabil-
ity) $+ \beta_5$ (parent size) $+ \beta_6$ (parent diversity) $+ \beta_7$ (parent relatedness) $+ \varepsilon$.

Variables

Industry growth. Industry growth is widely assumed to affect industry attractiveness favora-
bly, because it allows firms to grow without having to capture 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 & 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
industry growth is derived by estimating the growth in percent of industry sales between 1981
and 1985, as reported in Trinet.

Industry concentration. Industry concentration is usually assumed to have a positive rela-
tionship with industry profitability (Bain, 1956; Porter, 1980); scale economies and other sources
of market power presumably reduce the threat from potential entrants, allowing incumbents more

room to raise prices without inviting entry. Moreover, incumbents in concentrated markets are more likely to retaliate against potential entrants (e.g., by reciprocal entry). These considerations suggest a negative relationship between industry concentration and the probability of entry. The variable *industry concentration* is derived by estimating the 4-firm concentration ratio of each industry for 1981, based on Trinet Data.

Industry profitability. The relationship between entry and industry profitability is ambiguous. High profitability invites entry, but profitability is more likely when entry barriers are high (Baumol, Panzar and Willig, 1982); the net effect could go either way. To control for industry profitability we calculate a measure of the median return on assets (ROA) for each industry over the 1980–82 period. To calculate profitability we use all single-business firms in the Compustat industrial file for each 4-digit SIC industry, along with corresponding segments in the Compustat business segment file. Because of incomplete asset allocation, the profitability ratios in the segment file are systematically higher than the profitability ratios in the corporate database, so we standardize the observations from each database by computing them as percentage deviations from the database mean.[12]

Parent size. Parent size proxies for the parent's financial, managerial, and other resources. For a given level of relatedness, a large firm may be more willing and able to attempt entry, so we expect a positive relationship between parent size and the probability of entry. *Parent size* is measured as the total sales of the parent in 1981, based on Trinet data.

Parent diversity. The parent's level of diversification provides information about its strategy and motives. Firms active in many SIC codes may be following a strategy of broad or unrelated diversification, either to reduce risk, exploit internal capital markets, or leverage general re-

---

[12] Following Berger and Ofek (1995), we define industries at the 4-digit SIC level where five or more observations are available at that level, at the 3-digit SIC level if five or more observations are available only at that level, and at the 2-digit SIC level if five or more observations are available only at that level.

sources such as the ability to manage complex organizations. Such firms are more likely to enter additional industries, ceteris paribus. We thus expect a positive relationship between *parent diversity* and the probability of entry for any level of relatedness. *Parent diversity* is measured as the number of SIC codes for the parent in 1981, based on Trinet data.

Relatedness. Hypothesis 1 and 2 employ two measures of relatededness, $SURVTOT_i$ and $SICTOT_i$, both have been presented above, but we reiterate that they capture the sales-weighted average relatedness of the target industry i to *all other* businesses in the parent k. They are based on the survivor-based and SIC-based approach respectively.

Dependent variable. Our dependent variable is an indicator variable set equal to one if the parent entered a new 4-digit SIC code between 1981 and 1985 and zero otherwise. The Trinet database is used to identify entries and non-entries.

Variable definitions and data sources 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]

Results

The results from three logistic regressions are presented in Table 3. 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 $SICTOT_i$. Hypothesis 1, taking the idiosyncrasy interpretation as the point of departure, predicted that the coefficient on $SURVTOT_i$ would be insignificant, and that model 2 would not explain the probability of entry significantly better than model 1.

[Table 3 about here]

As shown in Table 3, H1 is not supported by the data. The coefficient on $SURVTOT_i$ is positive and highly significant, with a higher Wald score than any of the other independent variables (model 2). Moreover, the ability to explain the probability of entry increases substantially when $SURVTOT_i$ is added to a model with control variables only (moving from model 1 to model 2). The model χ^2 improves from 50.56 to 744.1, an improvement that is statistically significant beyond the 0.001 level. The two pseudo-R^2 measures increase from 2.1% to 26.9% (Cox and Snell R^2) and 2.8% to 35.8% (Nagelkerke R^2). The ability to predict entries and non-entries also increases from 55.3% to 74.4%. To the extent that the revealed preferences of the decision makers in our sample reflect information about relatedness, we conclude that relatedness does not appear to be mainly a firm specific variable. Decision makers in different firms appear to reach strikingly similar conclusion about the attractiveness of different combinations.

We now turn to H2, which takes the measurement interpretation as the point of departure. This hypothesis predicts that model 2 explains the probability of entry significantly better than model 3. As described above, the improvement in explanation can be interpreted as a measure of the magnitude of the measurement problems with the conventional SIC-based procedure. This, of course, requires that the explanatory power of the survivor-based measure $SURVTOT_i$ is not inflated for reasons unrelated to relatedness (a point we revisit below). Leaving this reservation temporarily aside, our revealed preferences data provide strong support for H2. All measures of model performance improve substantially when $SURVTOT_i$ is substituted for $SICTOT_i$. The model χ^2 improves from 227.26 to 744.1, an improvement that is statistically 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 12.2% to 35.8% (Nagelkerke R^2). These increases represent improvements of approximately 195%. The ability to predict entries and non-entries also increases from 61.4% to 74.4%, an improvement of 119.2%. The SIC-based approach is not worthless in terms of explaining entry de-

cisions, but it seems to capture a rather small share of the relatedness construct decision makers are acting upon. For other dependent variables and other samples, the portion captured by the SIC-measures may well be too small too allow significant and stable results. In sum we interpret this as providing preliminary support for the measurement interpretation.

TEST 2: PERFORMANCE

Hypotheses

An important objection to the revealed preference data in test 1 is that relatedness should be evaluated with respect to performance, not decision making. One may argue from an idiosyncrasy position that it is not the choice of combinations that are idiosyncratic, it is the performance obtained from them. If so, for the same reasons supporting H1, we would not expect $SURVTOT_i$ to contain much information about performance.[13] The performance metric we use here is survival. We assume that the worse the post entry performance, the more likely that an entered business will be exited (i.e. the entry decision is reversed). This gives us:

H3: $SURVTOT_i$ is a poor predictor of post entry survival

In a similar fashion one may argue that the question of measurement problems is better evaluated against a performance metric than decision making. Revealed preferences from observing decisions, reflects, at best, intentions and beliefs about performance outcomes. Differences in the ability to explain performance will therefore add to the credibility of the measurement hypothe-

---

[13] $SURVTOT_i$ is built on a procedure that extracts information about relatedness from other firms' actions, therefore this measure should not be able to pick up the firm specific components of relatedness.

sis. Using the probability of survival as our performance metric, the measurement interpretation yields:

H4: *SURVTOT$_i$* explains the probability of survival significantly better than *SICTOT$_i$*

Methods Test 2

We test Hypotheses 3 and 4 using a subsample of the sample used to test Hypotheses 1 and 2. We cannot use the full sample of 1,202 entries for two reasons. First, 1987 is the last year for which we have Trinet data on industry participation, needed to determine if a firm is still active in an entered industry. Because the competitive process may take time to filter entry decisions, we restrict the sample to entries made between 1981 and 1983. This reduces the sample to 401 cases. Using data from 1987 to measure post-entry survival poses a further challenge. The SIC system was revised in 1987, and there is no unambiguous procedure for converting between the two versions of the SIC system. Fortunately, 609 of the 913 relevant 4-digit SIC codes[14] were unchanged by the revision, and we restrict the sample to entries into these 609 industries. This further reduces our sample to 229 entries by 82 parents, all occurring between 1981 and 1983, for which their continued existence in 1987 could be unambiguously determined. Of these 229 entries 75 were exited by 1987.

We again use a logistic regression to test H3 and H4. The dependent variable is the probability that an industry entered between 1981 and 1983 is exited by 1987. The general model is the following: P (exit=1) = β_1 + β_2 (relatedness) + ε. The independent variables are the same two relatedness measures used to test H1 and H2. We do not include control variables in these regressions, since our main interest is the relative performance of the different relatedness measures.

---

[14] SIC codes referring to public and non-profit industries are omitted.

However, as a robustness check, we repeated the analysis reported below using controls for industry growth, industry concentration, industry profitability, parent size, and parent diversity. The substantive findings remained unchanged.[15]

Results

Table 4 presents the results of two logistic regression analyses. The column under the heading *SURVTOT* shows the regression output with $SURVTOT_i$. as the independent variable, while the column under the heading *SICTOT* shows the regression output when $SICTOT_i$ is the independent variable. H3, taking the idiosyncrasy interpretation as the point of departure, predicts that the coefficient on $SURVTOT_i$ will not be significant, and consequently, that the model with this variable included will poorly explain the probability of survival. As can be seen by table 4, this hypothesis is not supported by the data. The coefficient on $SURVTOT_i$ is negative (i.e. *SURVTOT* decreases the probability of exit) and significant at the 0.001 level. The model χ^2 is 23.47, also statistically significant at the 0.001 level. The two pseudo-R^2 measures are 9,7% (Cox and Snell R^2) and 13.6% (Nagelkerke R^2). We interpret this as additional evidence against the idiosyncrasy interpretation, since there is a component of relatedness that is not firm specific and is valuable not only for predicting entry decision, but also for predicting performance outcomes.

Hypothesis 4 is based on the measurement interpretation, predicting that the model containing $SURVTOT_i$ explains the probability of survival significantly better than the model containing $SICTOT_i$. As can be seen from table 4, H4 is supported. Replacing $SICTOT_i$ with $SURVTOT_i$ increases the model χ^2 from 5.90 to 23.47 and the two pseudo-R^2 measures increase from 2.5% to 9.7% (Cox and Snell R^2) and from 3.5% to 13.6% (Nagelkerke R^2). While the model with $SICTOT_i$ is not without any explanatory power, these increases in model performance indicate

---

[15] Regressions with control variables included are available from the authors.

that the SIC-based measure has a relatively poor ability to capture the aspects of relatedness that are important for performance outcomes. The relationship between the two measures is further explored in table 5. This table shows that adding $SURVTOT_i$ to the model containing $SICTOT_i$ as a regressor, the coefficient on $SICTOT_i$ is no longer statistically significant, while $SICTOT_i$ does not add any explanatory power to the model including only $SURVTOT_i$. In other words, the measure $SURVTOT_i$ seems to contain all the explanatory power of the measure $SICTOT_i$, and it also adds substantial explanatory power not contained in this measure. The two measures are therefore substitutes, not complements, in terms of capturing relatedness.

In summary, our data thus far are not consistent with the idiosyncrasy interpretation, while they fit a measurement interpretation well. We therefore put the idiosyncrasy interpretation aside, and turn our focus to the skeptic interpretation. The essential question arising from the skeptic interpretation is whether the results presented this far can be driven by other mechanisms than efficiency. In particular, we should be wary of mechanisms that could spuriously inflate the performance of the survivor-based measures versus the SIC-based measures, because such mechanisms could invalidate our interpretation of the results from the first two tests.

THE SKEPTIC INTERPRETATION VERSION I: HERD BEHAVIOR

The entry decisions we observe may reflect herd behavior rather than efficiency. "Rational" herding occurs when decision-makers suppress their private information, either because making a bad decision is less costly when others make the same decision (Scharfstein & Stein, 1990) or because decision makers believe that the decisions of others reflect valuable private information that other decision makers have (Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992). Either way, entry decisions may be based on the actions of others, rather than superior private knowledge about which activities are related to each other. Our survivor-based measures of relat-

edness are particularly susceptible to contamination from herd behavior, as they measure what combinations firms in the same industries as the focal firm have chosen previously. SIC-based approaches, on the other hand, are based on standardized distances in the SIC system and are little influenced by the behavior of other firms.[16] Support for Hypotheses 1 and 2 may therefore be driven by herd behavior, rather than a superior ability to capture relatedness.

However, a herd behavior interpretation is not consistent with the findings from the performance data in test 2. If the survivor-based measure primarily reflects herding, rather than efficiency, we would expect it to be a good predictor of entry but a poor predictor of post-entry performance (i.e. survival). Our underlying assumption here is that once entry has occurred, competitive forces will set in and begin screening the good decisions from the bad. The worse the post-entry performance, the more likely that an entered business will be exited. As our findings regarding Hypotheses 3 and 4 show, relatedness is indeed useful for predicting survival, and the survivor-based measure performs better than its SIC-based equivalent. The performance of the survivor-based measure is therefore difficult to dismiss as merely a better way to capture herding tendencies.

A possible counterargument is that exit decisions may also reflect herding, contaminating our measure of post-entry performance. We think this concern is misplaced, however. It implies that those firms least sensitive to herding before entry (those that chose rare combinations) become the most sensitive to herding after entry. The reversal of a previous entry decision seems more likely to be the result of performance below expectations than a sudden change of strategy. The herding version of the skeptic interpretation is therefore not able to convince us to change our main conclusion from tests 1 and 2, namely that a measurement interpretation fits the data best.

---

[16] Many industries that are close in the SIC system are never actually combined, while many seemingly distant industries are frequently combined.

THE SKEPTIC INTERPRETATION VERSION II: MUTUAL FORBEARANCE

The mutual forbearance hypothesis suggests that contact between firms across markets induces implicit or explicit agreements to refrain from aggressive competition (Edwards, 1955). Contact between firms across multiple markets enables a firm to respond to an aggressive action by a multipoint rival in one market with a retaliatory action in another market. Such behavior increases the potential costs of aggressive moves, leading to a less vigorous state of competition than what would have occurred without multi-market contact (Karnani and Wernerfelt, 1985). Firms may therefore prefer to enter industries where they will meet existing competitors as a mean of establishing mutual forbearance (Hypotheses 1 and 2), or they may refrain from exiting a weak position in one industry (Hypothesis 3 and 4) not because they reap efficiency gains from relatedness, but because they must remain in that industry to dissuade rivals from acting aggressively in other industries. There is some empirical support for the claim that the creation and exploitation of mutual forbearance affects the behavior and patterns of diversification (Greve and Baum, 2001).

This scenario is particularly worrying for two reasons. First, survivor-based measures of relatedness are particularly prone to incorporate motives of mutual forbearance because they are built by counting the frequencies of multi-market contact across industries. SIC-based measures, on the other hand, are probably less sensitive to such motives, because they are not constructed on the basis of counts of multi-market contact. Second, unlike the herd-behavior version of the skeptic interpretation, the mutual forbearance version is consistent with the findings in both test 1 and 2 above. To explore this possibility we therefore submit a third and final test.

Test 3: Mutual Forbearance

If diversification patterns primarily reflect attempts to establish or protect gains from mutual forbearance, rather than increase efficiency, we expect relatedness to be most effective at predicting entry into highly concentrated industries. The reason is that the mutual forbearance motive requires a minimum level of concentration to be plausible. In fragmented markets, existing or potential competitors without other markets to protect can force multi-market firms into vigorous competition. By the same reasoning, if the superior performance of the survivor based measure is due to contamination from mutual forbearance, we would expect the superiority of the survivor-based measure hold for entry into fairly concentrated industries, but no difference between the two measures with respect to the most fragmented markets. The mutual forbearance story therefore gives the following hypotheses:

H5: $SURVTOT_i$ will be a better predictor of entry when industry concentration in the destination industry is high, than when it is low.

H6: $SURVTOT_i$ will outperform $SICTOT_i$ in terms of predicting entry when concentration in the destination industry is high, but not when it is low.

Methods Test 3

To test Hypotheses 5 and 6 we use the same sample of 1,202 entries and 1,176 randomly chosen non-entries between 1981 and 1985 used to test Hypotheses 1 and 2. We divide the sample into two equally sized subsamples, one containing the most concentrated industries and one containing the least concentrated industries, and re-run on each sample the logistic regressions used to test Hypotheses 1 and 2. As before, the general model for the two sets of logistic regression

analyses is P (entry $= 1) = \beta_1 + \beta_2$ (industry growth) $+ \beta_3$ (industry concentration) $+ \beta_4$ (industry profitability) $+ \beta_5$ (parent size) $+ \beta_6$ (parent diversity) $+ \beta_7$ (parent relatedness) $+ \varepsilon$. The dependent variable is an indicator variable set equal to one if the parent entered a new 4-digit SIC code between 1981 and 1985 and zero otherwise.

Results and Discussion

Table 6 presents regression results for both the high and low concentration subsamples. Hypothesis 5 predicts that the model with controls and $SURVTOT_i$ will perform better for the high-concentration subsample. As can be seen from Table 6 this hypothesis is not supported. In fact the model with $SURVTOT_i$ performed better for the low-concentration subsample both in terms of the size of the coefficient (0.37 versus 0.43) and in terms of the increase in model χ^2 compared to a model with controls only (327,06 versus 365,31). The latter difference is statistically significant on the 0.01 level.

[Insert Table 6 about here]

Hypothesis 6 predicts that the relative superiority of the model with $SURVTOT_i$ over the model with $SICTOT_i$. will hold for the high-concentration sample, but not the low-concentration sample. The results in table 6 contradict Hypothesis 6. Not only is the survivor-based measure superior to the SIC-based measure in both subsamples, but the superiority of the survivor-based measure is bigger in the low-concentration subsample. This is seen most clearly from the changes in model χ^2 from model 2 to model 3. The changes in model χ^2 are not only positive and highly significant for both subsamples, they are actually larger for the low-concentration subsample than for the high-concentration subsample. In other words, the advantage of substituting survivor-based measures for SIC-based measures is greatest where the mutual forbearance interpretation is

least plausible. Overall, this leads us to reject the mutual forbearance version of the skeptic interpretation as a major influence on firms' diversification decisions, and also that contamination from mutual forbearance can explain the superior performance of the survivor based measure. We therefore uphold our previous interpretation that the superiority of the survivor based measure $SURVTOT_i$ is an indication of serious measurement problems with the conventional SIC-based approach to measuring relatedness.

CONCLUSIONS

This study is an attempt to shed empirical light on possible explanations for the mixed and unstable findings regarding a core hypothesis of the RBV; that related diversification is more efficient than unrelated diversification. We have isolated three candidate explanations. One is a measurement interpretation, arguing that inter-industry relatedness exists and is important, but has been poorly captured in much of the existing work. Another is the idiosyncrasy interpretation, which holds that the bulk of the empirical work does not take the RBV far enough in the sense of recognizing firm heterogeneity. Under this view, very little explanatory power should be expected when relatedness is studied as a general property of the relationship between industries. The third and final candidate is the skeptic interpretation which holds that the RBV's emphasis on (boundedly) rational pursuit of efficiency gains is misplaced. Other motives and/or other behavioral assumptions are needed to develop an empirically successful theory.

Our results suggests, in contrast to the skeptic interpretation, that there are significant and stable efficiency effects of relatedness, and that decision makers do act in a manner consistent with the pursuit of such effects. Contrary to the idiosyncrasy interpretation, these efficiency effects are not entirely firm specific. Thus it seems meaningful to speak of inter-industry relatedness. Moreover, we have documented the superior performance of survivor-based measures over

SIC-based measures. On the whole, we argue that measurement problems explain the mixed and unstable findings in the empirical literature using relatedness as an independent variable. While SIC-based measures do contain some information about relatedness, they seem to lack the sensitivity to detect anything but the very strongest relationships. The diversification decision is apparently a case in which the relationship is sufficiently strong for SIC-based measures to find it. The entry reversal decision, on the other hand, is more subtle, such that SIC-based measures may or may not detect a relationship, depending on which SIC-based measures one uses, sample size, sample composition, etc. Elsewhere we document a positive relationship between relatedness and bidder returns in mergers and acquisitions when a survivor-based measure is used, but no detectable relationship when SIC-based measures are used (reference withheld). This leads us to suspect that much of the confusion surrounding the role of relatedness will vanish if measures better capable of screening for the relevant theoretical conditions are employed. In sum, while the conventional RBV of diversification has been supported by our data, the most common operationalization of the RBV has been questioned.

The paper can also be read as an evaluation of the survivor-based approach to measuring relatedness. This approach allows the actions of firms in competitive markets to tell us (researchers) which industries are related to which, instead of imposing on the data some a priori view of relatedness. Rather than letting the SIC system or the researcher be the judge of what is related to what, we rely on the wisdom of local decision makers and the screening function of the competitive process. Of course, this procedure also means that we will not know the causes of relatedness in each particular instance, and hence will know little more about what relatedness *is*.[17] If this procedure is superior to conventional approaches, however, we may know much more about what

---

[17] 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. for example, t explain relatedness).

relatedness *does,* regarding its effects on dependent variables of interest (i.e. performance, growth, entry mode, financing, organizational choice, etc.).

To further validate the ability of survivor-based measures to capture relatedness it would be useful to complement the existing evidence with data from other periods and to link the survivor-based measures to behaviors and characteristics other than entry and entrant survival. One possibility is to see how well survivor-based measures explain additional performance measures such as revenue growth, return on assets, Tobin's *q,* or other measures of market value. While the research presented here uses the industry-pair as the unit of analysis, it is possible to compute average relatedness scores for individual firms as well.

In any case, the survivor-based approach does not (and should not) represent the end of the discussion about how to measure relatedness. Ideally, it would be preferable with measurement procedures that explicitly screen for the theoretical conditions associated with the concept (cf. the discussion in the second section above). However, as this seems infeasible at present, we must choose between imperfect alternatives. Our research suggests that the survivor-based approach may be a very strong contestant among these less-than-perfect alternatives.

REFRENCES

Alchian, A.A. 1950. Uncertainty, evolution, and economic theory. *Journal of Political Economy,* 58: 211–221.

Anderson E. 1985. The salesperson as outside agent or employee: A transaction cost analysis. *Marketing Science* 4: 234–254.

Arend, R.J. 2006. Tests of the resource-based view: do the empirics have any clothes? *Strategic Organization,* 4: 409–421.

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

Banerjee, A.V. 1992. A simple model of herd behavior. *Quarterly Journal of Economics,* 107(3): 797–817.

Barney, J. B. 1991. Firm resources and sustained competitive advantage. *Journal of Management,* 17(1): 99–120.

Baumol, W. J., Panzar, J. C., & Willig, R. D. 1982. *Contestable markets and the theory of industry structure.* New York: Harcourt.

Berger, P. G., & Ofek, E. 1995. Diversifications effect on firm value. *Journal of Financial Economics,* 37(1): 39–65.

Bigelow L. 2001. Efficient alignment and survival in the U.S. automobile industry, Working Paper, Olin School of Business, Washington University, St. Louis.

Bikhchandani, S., Hirshleifer, D., and Welch, I. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy,* 100(5): 992–1026.

Bulow, J. I., Geanakoplos, J. D., & Klemperer, P. D. 1985. Multimarket oligopoly: Strategic substitutes and complements. *Journal of Political Economy,* 93(3): 488–511.

Caves, R. E., Porter, M. E., & Spence, A. M. 1980. *Competition in the open economy: a model applied to Canada.* Cambridge: Harvard University Press.

Chatterjee, S., & Blocher, J. D. 1992. Measurement of firm diversification: Is it robust? *Academy of Management Journal,* 35(4): 874–888.

Chatterjee, S., & Wernerfelt, B. 1991. The link between resources and type of diversification: Theory and evidence. *Strategic Management Journal,* 12(1): 33–48.

Christensen, J. F., & Foss, N. J. 1997. Dynamic corporate coherence and competence based com-

petition: Theoretical foundations and strategic implications. In A. Heene & R. Sanchez (Eds.), *Competence-based strategic management.* Chichester: Wiley.

David, RJ, Han, S. 2004 A systematic assessment of the empirical support for transaction cost economics. *Strategic Management Journal* **25**: 39–58.

Dierickx, I., & Cool, K. 1989. Asset stock accumulation and sustainability of competitive advantage. *Management Science,* 35(12): 1504–1511.

Edwards, C. D. 1955. Conglomerate bigness as a source of power, *NBER conference report on business concentration and economic policy.* Princeton: Princeton University Press.

Eisenhardt KM. 1985. Control: organizational and economic approaches. *Management Science* 31: 134–149.

Elster, J. 1989. *Nuts and bolts for the social sciences.* New York: Cambridge University Press.

Fan, J. P. H., & Lang, L. H. P. 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.

Feinberg, R. M. 1985. Sales-at-Risk: A test of the mutual forbearance theory of conglomerate behavior. *Journal of Business,* 58(2): 225–241.

Finkelstein, S., & Haleblian, J. 2002. Understanding Acquisition Performance: The Role of Transfer Effects. *Organization Science,* 13(1): 36–47.

Foss, N. J., & Christensen, J. F. 2001. A market-process approach to corporate coherence. *Managerial and Decision Economics*, 22: 213–226.

Friedman, M. 1953. The Methodology of positive economics, In *Essays in positive economics.* Chicago: University of Chicago Press.

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, H. R, & Baum, J. A. C. 2001. Introduction: A multiunit, multimarket world, *Advances in Strategic Management*, Vol. 18: 1–28.

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

Hoskisson, R. E., & Hitt, M.A. 1990. Antecedents and performance outcomes of diversification: A review and critique of theoretical perspectives. *Journal of Management,* 16: 461–509.

Hoskisson, R. E., Hitt, M. A., Johnson,. R. A., & Moesel, D. D. 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.

Karnani, A., & Wernerfelt, B. 1985. Multiple point competition. *Strategic Management Journal,* 6(1): 87–96.

Klein, P. G. The make-or-buy decision: Lessons from empirical studies. In Claude Ménard and Mary Shirley, eds., *Handbook of New Institutional Economics.* Dordrecht: Springer, 2005.

Klein, P. G., & Klein, S. K. 2001. Do entrepreneurs make predictable mistakes? Evidence from corporate divestures. *Quarterly Journal of Austrian Economics,* 4: 3–23.

Lemelin, A. 1982. Relatedness in the patterns of industrial diversification. *Review of Economics and Statistics,* 64: 646–57.

Kwoka, J. E., & Ravenscraft, D. J. 1986. Cooperation v. rivalry: Price-cost margins by line of business. *Economica,* 53(211): 351–363.

Markides, C. C., & Williamson, P. J., 1994. Related diversification, core competence and corporate performance. *Strategic Management Journal,* 15: 149–165.

Markides, C. C., & Williamson, P. J. 1996. Corporate diversification and organizational structure: A resource-based view. *Academy of Management Journal,* 39: 340–367.

McFadden, D., & Manski, C. F. 1981. Alternative estimators and sample designs for discrete choice analysis. In D. McFadden & C. F. Manski (Eds.), *Structural analysis of discrete data with econometric applications*. Cambridge: MIT Press.

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

Montgomery, C. A., & 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*. Oxford: Blackwell.

Peteraf, M. A. 1993. The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal,* 14(3): 179–191.

Phillips, O. R., & Mason, C. F. 1996. Market regulation and multimarket rivalry. *Rand Journal of Economics,* 27(3): 596–617.

Porter, M. E. 1980. *Competitive strategy.* New York: Free Press.

Prahalad, C. K., & Bettis, R. A. 1986. The Dominant Logic: A New Linkage between Diversity and Performance. *Strategic Management Journal,* 7(6): 485–501.

Reed, R., & and Luffman, G. A. 1986. Diversification: the growing confusion. *Strategic Management Journal,* 7, 29–35.

Richardson, G. B. 1972. The organisation of industry. *Economic Journal,* 82(327): 883–896.

Robins, J., & Wiersema, M. F. 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, J. A., & Wiersema, M. F. 2003. The measurement of corporate portfolio strategy: Analysis of the content validity of related diversification indexes. *Strategic Management Journal,* 24(1): 39–59.

Rumelt, R. P. 1974. *Strategy, structure, and economic performance.* Cambridge: Harvard University Press.

Rumelt, R. P. 1982. Diversification strategy and profitability. *Strategic Management Journal,* 3(4): 359–369.

Salinger, M. A. 1984. Tobin's Q, unionization, and the concentration–profits relationship. *Rand Journal of Economics,* 15(2): 159–170.

Scharfstein, D. S., & Stein, J. C. 1990. Herd behavior and investment. *American Economic Review,* 80(3): 465–479.

Schmalensee, R. 1989. Inter-industry studies of structure and performance. In R. Schmalensee, & R. D. Willig (Eds.), *Handbook of industrial organization.* New York: North Holland.

Silverman, B. S. 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.

Silverman, B. S, Nickerson, J. A,, & Freeman, J. B. 1997. Profitability, transactional alignment, and organizational mortality in the U.S. trucking industry, *Strategic Management Journal,* 18: 31–52.

Shelanski, H. A., & Klein, P. G. 1995. Empirical Research in Transaction Cost Economics: A Review and Assessment. *Journal of Law Economics & Organization* 11: 335–361.

Stimpert, J. L., and Duhaime, I.M. 1997. In the eyes of the beholder: Conceptualizations of relatedness held by the managers of large diversified firms. *Strategic Management Journal,* 18(2): 111–125

Teece, D.J., Rumelt, R., Dosi, G., & Winter, S. 1994. Understanding corporate coherence: Theory and evidence. *Journal of Economic Behavior & Organization,* 23(1): 1–30.

Teece, D. J. 1980. Economies of scope and the scope of the enterprise. *Journal of economic behavior & organization,* 1(3): 223–247.

Teece, D. J. 1982. Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior & Organization,* 3(1): 39–63.

Voigt, C. W. 1993. *The Trinet large establishment data.* Unpublished manuscript. UCLA: Anderson Graduate School of Management.

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

Winter, S. G. 1971. Satisficing, selection, and the innovating remnant. *Quarterly Journal of Economics,* 85(2): 237–261.

Zenger TR, Marshall CR. 2000. Determinants of incentive intensity in group-based rewards. *Academy of Management Journal,* 43: 149–163.

TABLE 1

Variable Definitions and Data Sources

Variable	Definition	Data Source
Industry Growth	Salesgrowth 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
Parent Size	Total sales of the parent in 1981	Trinet
Parent Diversity	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

TABLE 2[a]

Means, Standard Deviations, and Correlation Coefficients of Independent Variables[a]

	Variables	Mean	SD	Industry Growth	Industry concentr.	Industry Profitab.	Parent size	Parent diversity	SURVTOT
	Industry variables								
1	Industry growth	0.60	1.64						
2	Industry concentration	32.21	21.42	0.15**					
3	Industry profitability	0.09	0.46	0.01	−0.04**				
	Firm variables								
5	Parent size	48727	56790	0.00	0.05**	−0.04*			
6	Parent diversity	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**

[a]$n = 2378$
$^{\dagger}p < 0.1$
$*p < 0.05$
$**p < 0.01$

TABLE 3[a]

Probability of Entry

| | Model 1 | | Model 2 | | Model 3 | |
| | Ctrls Only | | Ctrls + Survtot | | Ctrls + Sictot | |
	Beta	*Wald*	*Beta*	*Wald*	*Beta*	*Wald*
Constant	0.39**	13.82	−0.84*	41.71	0.16[†]	2.17
Industry growth	0.05*	3.99	0.06**	4.06	0.06*	4.90
Industry concentration	−0.01**	47.65	−0.01**	31.88	−0.02**	53.51
Industry profitability	−0.00	0.00	−0.13	1.43	−0.10	1.21
Parent size	0.00	0.00	0.00*	4.85	0.00	0.63
Parent diversity	0.00	0.38	0.00	0.33	0.00	0.64
$SURVTOT_i$			0.40**	403.33		
$SICTOT_i$					4.40**	99.94

[a]$n = 2378$
[†]$p < 0.1$
*$p < 0.05$
**$p < 0.01$

TABLE 3 (Cont.) [a]

Probability of Entry

	Model 1	*Model 2*	*Model 3*
	Ctrls Only	*Ctrls + Survtot*	*Ctrls + Sictot*
−2Log likelihood	3245.76	2552.23	3069.06
Model χ^2	50.56**	744.09**	227.26**
$\Delta \chi^2$ vs. model 1		693.53**	176.70**
$\Delta \chi^2$ model 2 vs. model 3		516.83**	
Cox and Snell R^2	0.021	0.269	0.091
Nagelkerke R^2	0.028	0.358	0.122
Correct classifications	55.3%	74.4%	61.4%

[a]n = 2378
[†]$p < 0.1$
*$p < 0.05$
**$p < 0.01$

TABLE 4

Probability of Exit

	SURVTOT	SICTOT
Variable coefficient	−0.16**	−1.65*
Wald statistic	16.64	4.31
Model χ^2	23.47**	5.90**
−2 log likelihood	266.17	283.74
Cox and Snell R^2	0.097	0.025
Nagelkerke R^2	0.136	0.035

[a]$n = 229$
[†]$p < 0.1$
*$p < 0.05$
**$p < 0.01$

TABLE 5 [a]

Models Containing Only SURVTOT, Only SICTOT, and Both

	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 χ^2	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	

[a] $n = 229$
[†] $p < 0.1$
[*] $p < 0.05$
[**] $p < 0.01$

TABLE 6 [a]

Probability of Entry

Model	1. Controls only		2. Ctrs +survtot		3. Ctrs + sictot	
Subsample (con-	High conc	Low conc	High conc	Low conc	High conc	Low conc
	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)
Industry growth	0.04 (1.24)	0.12* (4.87)	0.04 (1.19)	0.12* (4.56)	0.04 (1.37)	0.13** (6.33)
Industry concentration	−0.02** (34.03)	−0.01 (2.01)	−0.02** (17.18)	−0.03** (6.63)	−0.02** (40.50)	−0.02[†] (2.80)
Industry profit-ability	−0.06 (0.20)	0.08 (0.36)	−0.20 (1.99)	−0.02 (0.02)	−0.21[†] (2.63)	0.03 (0.05)
Parent size	0.00 (0.01)	0.00 (0.01)	0.00 (0.93)	0.00* (4.78)	0.00 (0.45)	0.00 (0.15)
Parent diversity	0.00 (0.01)	0.00 (1.03)	0.00 (0.09)	0.00 (0.27)	0.00 (0.00)	0.00 (1.46)
SURVTOT			0.37** (193.90)	0.43** (209.16)		
SICTOT					4.59** (56.35)	4.35** (45.92)
SURVNBOR						
SICNBOR						

[a]n =1189 (subsample with high concentration)
n = 1189 (subsample with low concentration)
[†]$p < 0.1$
*$p < 0.05$
**$p < 0.01$

TABLE 6 (Cont.)[a]

Probability of Entry

Model	1. Controls only		2. Ctrs +survtot		3. Ctrs + sictot	
Subsample (concen-	High conc	Low conc	High conc	Low conc	High conc	Low conc
	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)	Beta(Wald)
−2Log likelihood	1606.23	1630.52	1278.96	1265.20	1504.13	1551.13
Model χ^2	36.48**	7.78	363.54**	373.10**	138.34**	87.17**
Cox and Snell R^2	0.03	0.007	0.263	0.269	0.110	0.071
Nagelkerke R^2	0.04	0.009	0.352	0.360	0.147	0.095
$\Delta \chi^2$ vs. controls only			327.06**	365.31**	101.89**	79.38**
$\Delta \chi^2$ model 2 vs. model 3			High concentration sample: 225.20** Low concentration sample: 285.93**			

[a]n =1189 (subsample with high concentration)
n = 1189 (subsample with low concentration)
[†]p < 0.1
*p < 0.05
**p < 0.01