# WHY THERE? A DECOMPOSITION OF THE CHOICE OF TARGET INDUSTRY

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

*and*

Norwegian School of Economics and
Business Administration
Breiviksveien 40
N-5045 Bergen
Norway

*This version: 15.01 2008*

**Abstract**

This paper provides a quantitative quantitative characterization of how diversifying firms chose target industry, with an emphasis on target market and resource-/capability characteristics. We seek to improve on the existing literature by using two population level samples instead of the more restricted samples others have used, by removing important restrictions on the types of relatedness that can be captured, and by including measures of resource strength in addition to measures of relatedness.

**WHY THERE? A DECOMPOSITION OF THE CHOICE OF TARGET INDUSTRY**


The literature on corporate diversification has in recent years had its primary focus on the possible existence of a diversification discount, and whether the reported discount disappears when controlling for the endogeneity of the diversification decision (Berger and Ofek, 1995; Campa and Keida, 2002; Lang and Stulz, 1994; Maksimovic and Phillips, 2002; Santalo and Becerra, 2008; Villalonga, 2004a, 2004b).  This question is of course of crucial importance, but regardless of whether there is a diversification discount, the firms that do decide to diversify (or to diversify further) still have to decide which industries to target. This paper addresses this latter issue;  i.e. given that a firm decides to diversify, what determines the choice of target industry? This question received considerable attention in the 1980s and 90s, but much less so in the past decade (e.g. Chatterjee and Wernerfelt, 1991; Farjoun, 1994; Lemelin, 1982; Montgomery and Hariharan, 1991; Silverman 1999).

What we aim to bring to the table here is an examination of the broad patterns in the choice of target industry, in other words something similar in spirit to what the variance decomposition studies have done for corporate performance (McGahan and Porter1997; Rumelt, 1991; Short, Ketchen, Palmer and Hult, 2007).  Our intended contribution is therefore not new theory, but an examination of the relative importance and tradeoffs among variables identified in the received literature. Specifically, while most strategy researchers would presumably argue that industry attractiveness (Porter, 1980, 1985, 1987) and variables describing the relevance and strength of a firms preexisting resources and capabilities (Chatterjee and Wernerfelt, 1991; Montgomery and Wernerfelt, 1988; Penrose, 1959; Rumelt, 1974) are the most important factors in the choice of target industry, we do not know of any population level studies that can quantify how the probability of entry changes for a  given change  in such variables, how decision makers are

willing to trade off the different variables, and how much the different variables contribute to explain those decisions diversifying firms actually make. Existing work is limited either by sampling from a particular sector, such as manufacturing or publicly listed firms only (Chatterjee and Wernerfelt. 1991; Silverman, 1999), by restricting attention to a particular type of relatedness for example technological resources, human resource profiles or input ratios (Farjoun, 1994; Lemelin, 1982; Montgomery and Hariharan, 1992; Silverman, 1999) or from restrictions that arise from using SIC-distances as a proxy for relatedness (Chatterjee and Wenererfelt, 1991). We include all industries, and use a measure of relatedness designed to capture all kinds of relatedness.

We said above that we think of this as similar in spirit to the variance decomposition literature. One reason for using this analogy is that as in that literature we identify broad quantitative empirical patterns, and among them how much "variance" different variables explain. The quotation mark in the previous sentence refers to the fact in the present study we examine a dichotomous dependent variable, so we are using different statistical techniques that renders the analogy less than perfect. Another similarity is that as in that literature, an important underlying tension is the horserace between theory inspired by industrial organization and theory typically considered to belong to the resource based-/capabilities view. There are of course numerous other similarities and differences that will become clearer as we proceed.

To be able to characterize how firms choose target industries, and which variables explain those decisions, we rely on an unusually detailed population level dataset from the first half of the 1980s (AGSM Trinet). While somewhat dated, this sample has compensating advantages in its comprehensiveness and detail in terms of industry participation. Compared to Compustat, which only includes publicly listed firms, it provides full coverage of all firms (Voigt, 1993). Compustat is also based on SEC-filings, which requires separate reporting of all segments that

constitute 10% or more  of turnover. Trinet is a bottom up dataset, built from establishment level data. This means that Trinet provides a much more detailed breakdown of corporate portfolios than Compustat.

Another critical tool is that we use a fully continuous measure of inter-industry relatedness. This crucial variable is often measured using distances in the SIC or NACE systems. Since these classification systems are not truly continuous they impose statistical restrictions that do not apply to the survivor based approach we employ here. Equally important is that our procedure is flexible in the sense that it does not single out any particular source of relatedness to the exclusion of others. We capture relatedness by measuring how often different industries are performed together in the same firm, which means that we let local decision makers be the judge of what constitutes the relevant source of relatedness (Lien and Klein, 2008; Teece, Rumelt, Dosi and Winter, 1994)

In sum, we believe that our data provides a unique quantitative characterization of what explains the choice of target industry, and how firms weigh different variables. Our key findings are that relatedness is by far the most important variable in explaining where firms diversify. In our sample which consists of four industries a firm does not enter for each industry it does,  a one standard deviation increase in relatedness increases the probability of entry with somewhere between 25 and 29%. Next comes entry barriers, which - measured as industry concentration- decreases the probability of entry with about 5-6%. Other aspects of the target industry and firm resources and capabilities are statistically significant, but economically quite insignificant. In terms of explanatory power the dominance of relatedness is even more pronounced. Though almost all our variables are significant, about 85 % of the explanatory power of our full model comes from the relatedness variable.

**THEORY**

The industrial organization literature has documented beyond dispute that industries have different profitability levels (i.e. Bain, 1956; McGahan and Porter 1997; Schmalensee, 1989). This means that incumbency in some industries is more valuable than in others. But this should not, however, be taken to imply that IO-based reasoning suggests that firms should blindly rush to enter the industries with the highest growth or profitability levels. High return industries will tend to be associated with higher entry costs, and in long run equilibrium one would expect differences in return to fully reflect differences in entry costs, so that the expected returns from entering a high return industry will generally be no higher than the expected returns from entering a low return industry. Similarly, while high growth rates may be positive for incumbents - if entry barriers are high – high entry barriers implies that it is difficult for entrants to profit from high growth. If entry barriers are low, high growth tends to imply numerous entry attempts, depressing the survival chances of each entrant and possibly also the average return of the incumbents. General heuristics such as "enter a high growth/return industry" are therefore questionable, because such "rules for riches" cannot exist in equilibrium.

Nevertheless, empirical evidence strongly suggests that industries are generally not in this kind of entry equilibrium, and if they are, they are not likely to remain so for very long (Geroski, 1995; Baldwin, 1995). Outside equilibrium, it is precisely the tendency for firms to prefer entry into high growth, high profitability industries that creates a force tending towards equilibrium. If we grant that most industries are outside equilibrium most of the time, then variations in industry profitability and -growth should significantly affect entry behavior. Entry behavior is, of course, a matter of both new start-up firms, and diversification by existing firms. Our focus here is on diversification. Firms making diversification decisions are likely to perceive that they are faced

with several potential target industries. A key point in the IO literature, then, is that in choosing between these opportunities, firms will display a systematic bias in favor of opportunities in attractive target markets. In his 1980 classic, Michael Porter explicitly argued that firms should screen potential target markets for the strength of the competitive forces, and make this a key factor in entry decisions (Porter, 1980)[1].

There is considerable empirical support for the claim that target market conditions influences entry decisions. Entry is positively associated with industry growth, and negatively influenced by various types of entry barriers (Baldwin and Gorecki, 1987; Dunne and Roberts, 1991; Geroski, 1991; Kessides, 1990; Khemani and Shapiro, 1986, 1988; Mata, 1991, 1993; Salinger, 1984)[2]. What is somewhat less clear is the relationship between (historical) profit rates and entry. While Geroski (1995) concludes that entry responds only very slowly to profit differentials, Siegfried and Evans (1994) concludes that there is a clear positive link. The more ambiguous findings on this variable is not entirely surprising given that high profit levels may signal high entry costs as much as it signals high expected profits.

Below we shall use three variables to characterize target market attractiveness. One is industry growth, which we expect to increase the probability that a diversifying firm selects a given industry. The second is industry concentration, which we use as a summary proxy for entry barriers. We expect concentration to be negatively related to the probability that a given industry is targeted. The third is industry profitability. Industry profitability is not unrelated to growth and concentration, indeed they are conventionally seen as key causes of industry profitability. Industry profitability is therefore likely to be somewhat confounded by the former two variables. However industry profitability is not fully explained by growth and concentration, so we include

---

[1] An additional argument may be that firms will systematically attempt to develop capabilities that will allow them to enter more attractive markets. Thus target market attractiveness will steer capability development. Conversely, under the RBV, capability development drives the choice of markets.

[2] Note that this literature as a rule does not distinguish between de novo entry, and entry via diversification.

it partly to control for other mechanisms that influence profitability (e.g. substitutes, vertical bargaining power, etc.), and partly to because it is possible that decision makers use it as an important statistic in its own right. When we use industry profitability in regressions where entry barriers are controlled for by the industry concentration variable, we expect it to be positively associated with the likelihood of entry.

The resource based view (henceforth: RBV) takes a different starting point. With Edith Penrose's landmark work on the growth of firms as the point of departure, the RBV of diversification sees entry by diversified firms as driven by excess capacity in existing resources and capabilities (Penrose, 1959). The choice of target markets are driven primarily by considerations about relevance and strength of this excess capacity in various applications, while the attractiveness of the target market plays only a secondary role. So where you go depends on what you have got (in excess). Relevance is usually summarized by relatedness between a target industry and one (or more) of the diversifiers existing industries. There is a long line of studies that confirm the idea that some industries are more related than others, and that firms display a strong preference for diversification into those that are related to their own (e.g. Chatterjee and Wernerfelt, 1991; Farjoun, 1994; Lemelin, 1982; Montgomery and Hariharan, 1991; Silverman 1999)[3].

In line with the resource-based view one would assume that a diversifying firm will prefer target industries where their resources and capabilities are not only relevant, but also emphasize using their strongest resources as a basis for diversification. Oddly, the amount of research done on the impact of resource strength is dwarfed by the research on resource relevance (i.e. relatedness). This is particularly odd because the other main arm of the resource based view, the one that focuses on competitive advantage, is all about the importance of resource heterogeneity

---

[3] Note that while there is disagreement as to whether related diversifiers outperforms unrelated there is little doubt over the tendency of firms to prefer related diversification.

(Barney, 1986, 1991; Dierickx and Cool, 1989; Peteraf, 1993). Nevertheless, we feel safely within the confines of the resource based view when we expect to find that firms are more likely to target related industries, and more likely to target industries that involves diversification on the basis of strong resources. We should emphasize here that by strong we mean both strength relative to the other resources in the diversifiers portfolio, and strong relative to competitors. Both types of strength are expected to increase the probability of entry, as is relatedness.

## DATA AND METHODS

### Sample

The ideal sample for this work would be a population level sample of all diversification decisions made over some period of time. As mentioned above, the AGSM Trinet database is such a sample. Trinet contains  biannual records of all U.S. establishements with more than 20 employees, including variables such as 4-digit SIC code, corporate ownership and sales. It's been found to include 95 % of all the establishments it should, which, unlike Compustat, includes both listed and unlisted firms (Voigt, 1993). Trinet covers the period from 1979 through 1989, but only the data from 1981, 1983 and 1985 are usable for our purposes. The reason that the data from 1979 and 1989 are not usable has do with changes in coding practices in Trinet which makes identifying diversification moves uncertain . The 1987 data are also difficult to use because the SIC-classification system was changed this year, and there is no unambiguous way to convert data between the two versions. This makes it difficult to determine whether a firm makes a diversification move between 1985 and 1987, or whether observed changes are due to the SIC revision.

This leaves us with data from 1981, 1983 and 1985. We start with the database from 1981 and record all industries a given firm was active in i that year.  Next we compare this with the

1983 data, and register which new industries each firm had entered between 1981 and 1983. We include all diversification moves by all firms in our sample, making it a de facto population level record of diversification moves between 1981 and 1983. Note also that we *only* include firms actually making diversification moves. This reflects that our focus is not on why firms diversify, nor on whether such firms are different from other firms, but rather on where they go given the decision to diversify. Next we repeat the procedure with the data from 1983 and 1985, giving us a second sample, containing all diversification moves between 1983 and 1985. In addition to the actual diversification moves, for each diversification move by each firm, we randomly select four possible target industries that were not chosen by that firm (or already present in the firm's portfolio). So in essence, much of the empirical exercise we describe below involves predicting which of the five potential target industries is the one actually entered. We could of course have included all industries that were not entered instead of four. However, this would lead to a sample completely dominated by non entry. McFadden and Manski (1981) suggest using state based sampling in situations when a sample is overwhelmingly characterized by one state, and demonstrate that this will provide unbiased and consistent coefficients for all variables except the constant term.

The resulting 1981-1983 sample consists of 2592 firms, that entered 6377 new industries, and we added to this 25508 randomly chosen non entries (4 for each case of actual entry), but cases with missing data reduced this number to 24980. This means that this sample contains a total of 31357 cases. The 1983-1985 sample consists of 2440 firms, that entered 5849 industries. We added to this 23396 randomly chosen non entries, but missing data reduced this number to 23188. The resulting sample therefore contains a total of 29037 cases.

One drawback of the Trinet database is that it does not provide financial information. This means that in order to calculate industry profitability, we had to use financial data from

Compustat. This variable is therefore calculated from profitability levels of listed firms only. We also used Compustat data to calculate industry growth for 1980-1981, since this cannot be calculated using Trinet. So for the first sample our industry growth variable is calculated from publicly listed firms only. Industry growth for 1982-1983 is calculated using Trinet, so all firms are included in calculating industry growth for the second sample. To examine whether this is a problem we correlated sales measures between Trinet and Compustat for the period 1981-1985, the resulting correlation is 0.893, which indicates that this not a major problem (also, as seen below, the coefficients on industry growth is very similar across the two periods).

**Measures**

Our dependent variable is dichotomous. In each of the two samples we record whether a firm enters a potential target industry (entry = 1), or not (entry = 0). In the first sample this means that the firm enters (or not) after 1981, but before 1983. In the second sample it means that the firm enters (or not) after 1983, but before 1985.

Our independent variables includes two control variables, three variables that aim to capture industry attractiveness, and yet another three variables that seeks to capture the relevance and strength of the firms preexisting resources and capabilities.

The two control variables included are designed to control for the possibility that large firms may make decisions differently than smaller firms. We control for two aspects of firm size. The first, *parent sales* records total sales by the parent firm as reported in Trinet in the year immediately preceding the entry period (e.g. 1981 and 1983, respectively). The second is *parent diversity*, which records the number of industries participated in by the parent. This variable is also derived from Trinet in the year preceding the entry period.

The first of the industry attractiveness variables is *industry concentration*. We measure industry concentration as the C4 ratio of the target industry in the year preceding the entry period. Data from Trinet.

The second industry variable is *industry growth*. For industry growth we focus on the two years before the entry period. In other words, for entries between 1981 and 1983 we focus on industry growth in the period from 1980 to 1981 (note this involves no overlap with the entry period), similarly, for entries between 1983 and 1985, we focus on growth in the period 1982 to 1983. As stated above the measure of industry growth for the former period is calculated from Compustat data, while the latter is from Trinet data. In both instances industry growth is calculated as the percentage change in aggregated industry sales over the relevant years. For the Compustat data we aggregate data from the segment database as well as from non diversified firms in the firm level database.

The third industry level variable is *industry profitability*. Here too we focus on the two years preceding the entry period.  In calculating historical profitability levels for an industry there are several options. One can calculate the mean return of all incumbents, or in order to reduce the impact of outliers, one can calculate the median return. A third option is to aggregate all returns and divide this by all assets in the industry. Unlike the two former options, this provides a size weighted measure, or in other words the return on the average dollar invested in the industry. We experimented with all three options, but decided on the latter. While the three different options yielded materially similar findings, we chose the weighted version simply because it performed marginally better in terms of pseudo $R^2$ than using mean or median ROA.  The data for computing industry profitability were taken from Compustat. Compustat contains both a segment database and a firm level database, we use all the data from the segment database, but only non diversified firms from the firm level database.

We now turn to the variables characterizing the relevance and strength of the firms preexisting resources and capabilities. In terms of relevance, we measure the variable *relatedness* using the survivor based approach (Lien and Klein, 2008;  Teece et al., 1994). What this measure does is capture how related the closest preexisting industry in the parent firms portfolio is to the target industry. Put differently we obtain the minimum distance from any industry in the firms existing portfolio to the potential target industry[4]. To compute this we needed a measure of the relatedness between all industries in the economy. The survivor based approach used here involves estimating how often a given pair of industries is combined by firms, compared to the number of combinations one would expect if diversification patterns were random (adjusting for industry size). This difference is taken to reflect the degree of relatedness between a given pair of industries. I.e. industries are related to the extent that this difference is large and positive, and it is negative to the extent that the difference is negative. The procedure was originally developed by Teece et al. (1994), and have been evaluated as a measure of relatedness by Lien and Klein (2008), who found it to be substantively and consistently superior to the conventional method of relying on SIC-distance measures. An advantage of this measure over other types of relatedness measures is that it does not single out any particular type of relatedness such as technological resources (Jaffe, 1986; Robins and Wiersema, 1995; Silverman, 1999) or human resource profiles (Farjoun, 1994). In principle it captures all kinds of relatedness.  Details of how this measure is constructed can be found in appendix A. Note though that the data used to find the relatedness between all pairs of industries are all diversified firms in the US-economy for the relevant year. This data is subsequently used to identify the highest scoring pair between any of the existing industries in a firms portfolio, and the target industry. It is the score of this pair that

---

[4] We also experimented with including measures of how close the second closest industry is to the target industry. This variable is also highly significant and positively related to actual entry, but the coefficient is substantially smaller. When we do not report this, it is because including it means losing all observations of firms with only one preexisting industry. This would mean that we no longer have a population level study.

constitutes our variable *relatedness*. We use Trinet data for the year preceding the entry period to calculate this variable.

The next variable is based on the idea that firms would prefer to diversify on the basis of strong resources and capabilities. One measure of strength is whether the firm has above average market share in the closest related industry. We calculate the variable *share strength* by measuring the market share of the focal firm in that industry minus the mean market share of all firms in that industry, divided by the standard deviation of market share for firms in that industry. Therefore *share strength* measures, in standard deviation units, how far above or below average the firm's market share is. We do acknowledge that market share is not a perfect measure of resource strength, but there is no data source that can provide better performance data on such a disaggregated level. Also, and more importantly, high market share is consistently reported as positively associated with high economic performance (e.g. Gale, 1971; Gale and Branch, 1982; Shepard, 1972).  We therefore find this to be a noisy, but acceptable measure. This variable is also calculated using Trinet data for the year preceding the entry period.

The third and final variable complements the former. The logic is here that the firm will want to base expansion on the strongest of its in-house resources and capabilities. We take the position that the businesses that constitute the largest part of a firm's sales are the businesses with the strongest resources and capabilities. We therefore again look at the closest related existing industry to the potential target industry, and measure how big a share of a firms total sales is derived from this business. We use this share to define our final variable *intra strength*. The variable is calculated from Trinet data from the year preceding the entry period. The means, standard deviations and correlations of all variables are provided in table 1.

-------------------------------------------------------------------------

Insert table 1 about here

------------------------------------------------------------------------

**Statistical Approach**

In this study the dependent variable is dichotomous, either a potential target industry is entered, or it is not. This makes logistic regression appropriate. The general model is the following:

$$\text{Logit } Y = \alpha + \beta_1 \text{Parent Sales} + \beta_2 \text{Parent Diversity} + \beta_3 \text{Industry}$$
$$\text{Concentration} + \beta_4 \text{Industry Growth} + \beta_5 \text{Industry Profitability} + \quad (1)$$
$$\beta_6 \text{Relatedness} + \beta_7 \text{Share Strength} + \beta_8 \text{Intra Strength} + \varepsilon$$

Logit Y is the natural logarithm of the odds that an industry was actually entered :

$$\ln [P(Y=1) / (1-P(Y=1)] \quad (2)$$

We are interested in two things from equation (1). We are interested in the estimates of $\beta_3$ - $\beta_8$. Through these we can, after suitable transformations, characterize how each of them affects the probability of entry, or put differently; how decision makers weigh and trade off the different variables when choosing target industry . In addition, we are interested in the contribution from the different variables in terms of explaining the dependent variable, analogous to $R^2$ in linear regression. Care must be taken in interpreting both the variable coefficients and contributions to pseudo $R^2$ since logistic regression is not linear and the dependent variable is the logarithm of the odds of entry.

**FINDINGS**

Table 2 presents the outputs from the logistic regressions. We run three sets of models. The first set, model 1 and 2, includes only the two control variables. Neither is significant in either of the two samples. Next, we add the industry attractiveness variables. As we see from model 3 and 4 *industry concentration* is negatively signed as expected, while *industry growth* is positively signed. Industry growth has a much larger coefficient in the last sample, but this difference in coefficient size disappears when the resource based variables are added. What is most striking about model 3 and 4 is, however, the way *industry profitability* changes sign between the two samples. It is significant and positively signed in the 1981-83 sample, and significant and negatively signed in the 1983-85 sample. It is tempting to speculate that this has something to do with unrelated diversification going out of fashion during this period. On the other hand, this interpretation is somewhat contradicted by model 5 and 6. First, we see that the negative coefficient in the later sample loses significance when the resource based variables are added. Also, if there was a marked increase in how decision makers value relatedness, one would expect the coefficient on *relatedness* to be larger in model 6 than in model 5.  We actually find that it is slightly smaller. Finally, the negative coefficient on the control variable *parent diversity* in model 5 and 6 is also smaller in the later sample, in contrast to what the "unrelated is unfashionable" story would predict. Our conclusion is therefore that high historical profitability levels is a weak and unclear signal of attractiveness. This is corroborated further below when we examine the coefficient sizes and explanatory contribution of the variables in greater detail.

In general, in the full models (5 and 6) we can see that with the exception of *industry profitability*, all the independent variables are significant, and signed as expected. The coefficient sizes in model 5 and  6 are also quite stable across the two samples. All the resource based

variables are significant and signed as expected. What is also striking in table 2 is how the model performance change. The regression with only the control variable is not significant. As we enter the industry attractiveness variables the model becomes significant, with a pseudo $R^2$ of between 6 .5% and 5.5%. When we add the three resource based variables the pseudo $R^2$ jumps to between 40.8% and 35.6%. Evidently the latter group of variables add contribute the most to the explanatory power of the model.

-----------------------------------------------------

Insert table 2 about here

--------------------------------------------------

From table 2 it can be rather difficult to get a sense of what the coefficient sizes means for how decisions are made, and how decision makers are willing to trade off different variables. This is particularly difficult in logistic regression, because of the nonlinearity of the model. In table 3 we try to make the implications of the coefficient estimates more transparent. In table 3 we calculate how different changes in each of the independent variables change the probability that a potential target industry is actually entered by a diversifying firm. In particular we look at a 1 standard deviation (positive) change from the mean, and a 10% and a 100 % change from the mean. Note also that all other variables are held at their mean values when the effects of changes in a given variable is examined.
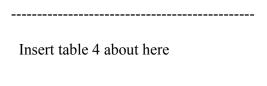
---------------------------------------------------------------------

Insert table 3 about here

---------------------------------------------------------------------

What can be seen from table 3 is that not all variables that are statistically significant are substantively significant. In fact, the variables *industry growth*, *industry profitability*, and *share strength* all have very little impact on the probability of entry. It would generally take an

enormous change in any of these variables to offset small changes in any of the remaining three variables, *industry concentration, relatedness* and *intra strength*.  In particular, relatedness is by far the variable that changes the probability of entry the most. A 1 standard deviation change in *relatedness* increases the probability of entry by somewhere between 25 and 29%. Next comes *industry concentration*. A 1 standard deviation increase in *industry concentration* reduces the probability of entry by between 5 and 6 %. The third substantively significant variable is *intra strength*. As can be recalled this variable measures how big a share of total firm sales that come from the closest related industry. A 1 standard deviation increase in this variable increases the probability of entry by 3.24%.

We now turn to examine the explanatory contribution from the individual variables in greater detail. Following Menard (2002), we do this by evaluating how the -2*Log Likelihood scores in model 5 and 6 from Table 2 can be decomposed into the contribution from the different variables[5]. This is provided in Table 4 below.

-----------------------------------------------

  Insert table 4 about here

-----------------------------------------------

As we see from table 4, in the neighborhood of 86-88 % of the reduction in -2*Log Likelihood comes from the relatedness variable, 7-9% from industry concentration, and all other variables contribute less than 2%. We therefore conclude that the relatedness variable completely dominates the explanatory power of models 5 and 6 in Table 2. The only other sizable contribution comes from industry concentration. The relevance of preexisting resources and capabilities together with entry conditions in the target market appears to be what explains entry decisions best.

---

[5] We use the backward conditional estimation procedure in SPSS to arrive at these estimates

In addition to the findings reported here, we have examined various interaction effects and nonlinear terms. Though this work is not completely finalized yet, we have not found anything that materially changes the results presented here.

## CONCLUSIONS AND CAVEATS

Our aim here has been to characterize the choice of target industry by diversifying firms, and to do so in quantitative terms. To do so we have examined a population level dataset that includes virtually all diversification move made in the US economy in the period 1981-1985, and for each actual diversification move, we randomly chose 4 potential target industries that were not entered. Our main finding is that the average decision maker is mostly concerned with finding destination industries were existing resources and capabilities are relevant, next on the list is to avoid industries that are difficult to enter. Also our findings provide some support that firms prefer to diversify on the basis of the largest existing businesses, which presumably contains the relatively stronger resources and capabilities in the firm's portfolio. The historic growth and profitability of the target market, and the market share performance of the firm in the closest related market, matters much less for decision makers. In terms of explanatory power, the dominance of *relatedness* is even more pronounced.

We were somewhat surprised that the variable *share strength* did not have a bigger impact. This could be attributable to measurement error. Market share performance is certainly not an ideal measure of resource strength. Another possibility is that share strength is only relevant when relatedness is high. We did examine this issue using interaction terms, but doing so hardly contributed any additional explanatory power. Yet another possibility is that many of the firms

making diversification decisions are trying to find new arenas to apply resources and capabilities that are performing poorly in their existing markets. This is consistent with the story provided by authors claiming that the diversification discount is sprurious, and arises because poorly performing firms are the ones most likely to diversify (Kampa and Keida, 2002; Maksimovic and Phillips, 2002; Villalonga, 2004a, 2004b). This might seem inconsistent with the point above about diversifying on the basis of the strongest internal resources and capabilities, but it might be the case that a poorly performing large business gives especially strong incentives to attempt diversification.

Nevertheless, we feel it is appropriate to end by pointing out that a large population level study such as this inevitably destroys complexity. While the findings reported here may be true on average and for the economy as a whole, there will inevitably be many instances involving specific firms and specific industries that depart from this picture. However, our goal here has been to provide the big picture.

**APPENDIX A**

The calculation of the survivor based measure of relatedness between a given pair of industries 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:

$$\Pr(X_{ij} = x) = f_{hg}(x, K, n_i, n_j) = \frac{\binom{n_i}{x}\binom{K - n_i}{n_j - x}}{\binom{K}{n_j}} \tag{3}$$

The mean and variance of $X_{ij}$ are:

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

(4)

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

(5)

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

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

(6)

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. Our variable *relatedness* is defined by identifying the highest score of $SR_{ij}$ that can be constructed between a potential target industry and any industry the firm in question is already active in. The $SR_{ij}$ score of that pair defines the variable *relatedness*.

**REFRENCES**

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

Baldwin, J. R. 1995. ***The dynamics of industrial competition***. Cambridge University Press: Cambridge.

Baldwin, J. R., & Gorecki, P.K. 1991. Entry, exit and productivity growth. In P.A. Geroski and J. Scwalbach, eds, ***Entry and market contestability***. Oxford: Basil Blackwell: 244-256.

Barney, J.B. 1986. Strategic Factor Markets - Expectations, Luck, and Business Strategy. **Management Science**, 32: 1230-1241.

Barney, J.B. 1991. Firm Resources and Sustained Competitive Advantage. ***Journal of Management,*** 17: 99-120.

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

Campa, J. M., & Kedia, S. 2002. Explaining the diversification discount. ***Journal of Finance,*** 57: 1731–1762.

Chatterjee, S. 1990. Excess Resources, Utilization Costs, and Mode of Entry. The ***Academy of Management Journal***, 33 (4): 780-800.

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

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

Dunne, T., & Roberts, M.J. 1991. Variation in producer turnover. In P.A. Geroski and J. Scwalbach, eds, ***Entry and market contestability***. Oxford: Basil Blackwell: 187-203.

Farjoun, M. 1994. Beyond industry boundaries: Human expertise, diversification and resource-related industry groups. *Organization Science,* 5(2): 185–199.

Gale, B.T. 1971. Market share and rate of return. *Review of Economics and Statistics* 54: 412-423.

Gale, B., & Branch, B. 1982. Concentration versus Market Share: Which Determines Performance and Why Does It Matter? *Antitrust Bulletin*, 27: 83-105.

Geroski, P.A. 1995. What do we know about entry? *International Journal of Industrial Organization,* 13: 421-440.

Jaffe, A.B. 1986. Technological opportunity and spillovers of R & D: Evidence from firms' patents, profits, and market value. *The American Economic Review*, 76 (5): 984-1001.

Kessides, I.N. 1990. Toward a testable modell of entry: A study of the US manufacturing Industries. *Economica*, 57:219-238.

Khemani, R.S., & Shapiro, D.M. 1986. The determinants of new plant entry in Canada. *Applied Economics*, 18: 1243-1257.

Lang, L.H.P, & Stulz; R.M. 1994. Tobin's q, diversification and firm performance. *Journal of Political Economy*, 102: 1248-1280.

Lemelin, A. 1982. Relatedness in the Patterns of Interindustry Diversification.*The Review of Economics and Statistics,* 64: 646-657.

Lien, L.B., & Klein P.G. 2008. Using Competition to Measure Relatedness. *Journal of Management.* published on December 22, 2008 as doi:10.1177/0149206308328505

Maksimovic, V., & Phillips, G. 2002. Do Conglomerate Firms Allocate Resources Inefficiently Across Industries? Theory and Evidence. *Journal of Finance,* 57: 721-767.

Mata, J. Sunk cost and entry by small and large plants. In P.A. Geroski and J. Scwalbach, eds, *Entry and market contestability*. Oxford: Basil Blackwell: 49-62.

Mata, J. 1993.Entry and type of entrant: Evidence from Portugal. *International Journal of Industrial Organization,* 11: 101-122.

Menard, S. 2002. Applied logistic regression analysis. Volume 106 of Sage University Papers. *Quantitative Applications in the social sciences*.

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.

McGahan, A., & Porter M.E. 1997. How much does industry matter, really? *Strategic Management Journal*, 18: 15-30.

Montgomery, C. A., & Hariharan, S. 1991. Diversified expansion by large established firms. *Journal of Economic Behavior & Organization,* 15(1): 71–89.

Montgomery, C.A., & Wernerfelt, B. 1988. Diversification, Ricardian Rents, and Tobin-Q. *Rand Journal of Economics,* 19: 623-632.

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: 179-191.

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

Porter, M. E. 1985. *Competitive advantage*. New York: Free Press.

Porter, M. E. 1987. From competitive advantage to corporate strategy. *Harvard Business Review,* 65(3): 43–59.

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.

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

Rumelt, R. P. 1991. How much does industry matter? *Strategic Management Journal*, 12 **:**167-
185.

Salinger, M.A. 1984. Tobin's q, Unionization, and the Concentration-Profits Relationship. *Rand
Journal of Economics*, 15: 159-170.

Santalo, J., & Becerra, M. 2008. Competition from specialized firms and the diversification-
performance linkage. *Journal of Finance*, 68 (2): 851- 883.

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.

Short, J.C., Ketchen, D.J., Palmer, T.B., & Hult, G.T.M. 2007. Firm, strategic group, and
industry influences on performance. *Strategic Management Journal,* 28: 147-167.

Sigfried, J.J., & Evans, L.B. 1994. Empirical studies of entry and exit: A survey of the evidence.
*Review of Industrial Organization*, 9: 121-155.

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.

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.

Villalonga, B. 2004a. Does diversification cause the diversification discount? *Financial
Management,* 33(2): 5–27.

Villalonga, B. 2004b. Diversification discount or premium? **Journal of Finance**, 59:479-505.

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

**Table 1: Means, Standard Deviations , and Correlation Coefficients of Independent Variables, 1981-83 Sample**

| | Mean | Standard Dev. | Parent Sales | Parent Diversity | Industry Conc. | Industry Growth | Industry Prof. | Relatedn. | Share Strength | Intra Strength |
|---|---|---|---|---|---|---|---|---|---|---|
| Parent Sales | 12350 | 28821 | 1 | | | | | | | |
| Parent Diversity | 17.11 | 20.47 | 0.557*** | 1 | | | | | | |
| Industry Conc. | 33.67 | 23.33 | 0.019*** | 0.020*** | 1 | | | | | |
| Industry Growth | 0.20 | 1.08 | 0.003 | -0.001 | 0.011** | 1 | | | | |
| Industry Profitability | 0.15 | 0.19 | 0.001 | 0.000 | 0.034*** | 0.021*** | 1 | | | |
| Relatedness | 8.95 | 9.84 | 0.139*** | 0.236*** | -0.108*** | 0.019*** | 0.002 | 1 | | |
| Share Strength | 0.04 | 0.95 | 0.164*** | 0.061*** | -0.005 | 0.002 | 0.006 | 0.051*** | 1 | |
| Intra Strength | 0.16 | 0.25 | -0.219*** | -0.369*** | -0.024*** | 0.002 | 0.013** | -0.043*** | 0.209*** | 1 |

***, **, and * represent statistical significance at the 1, 5, and 10 percent levels, respectively. N=31357 for the 1981-83 sample.

**Table 1 (Cont.): Means, Standard Deviations , and Correlation Coefficients of Independent Variables, 1983-85 Sample**

| | Mean | Standard Dev. | Parent Sales | Parent Diversity | Industry Conc. | Industry Growth | Industry Prof. | Relatedn. | Share Strength | Intra Strength |
|---|---|---|---|---|---|---|---|---|---|---|
| Parent Sales | 20548 | 50023 | 1 | | | | | | | |
| Parent Diversity | 19.60 | 23.97 | 0.477*** | 1 | | | | | | |
| Industry Conc. | 33.33 | 23.07 | 0.009 | 0.020*** | 1 | | | | | |
| Industry Growth | 0.32 | 0.54 | 0.000 | 0.001 | -0.012** | 1 | | | | |
| Industry Profitability | 0.17 | 0.78 | -0.003 | -0.004 | 0.058*** | -0.005 | 1 | | | |
| Relatedness | 9.47 | 10.54 | 0.137*** | 0.249*** | -0.105*** | 0.078*** | -0.016*** | 1 | | |
| Share Strength | 0.08 | 1.09 | 0.163*** | 0.027*** | -0.013** | -0.010 | 0.000 | 0.026*** | 1 | |
| Intra Strength | 0.15 | 0.25 | -0.203*** | -0.356*** | -0.029*** | 0.000 | -0.006 | -0.057*** | 0.233*** | 1 |

\*\*\*, \*\*, and \* represent statistical significance at the 1, 5, and 10 percent levels, respectively. N= 29037 for the 1983-85 sample.

**Tabel 2: Logistic Regression Output**

| Variable | Model 1 1981-83 | Model 2 1983-85 | Model 3 1981-83 | Model 4 1983-85 | Model 5 1981-83 | Model 6 1983-85 |
|---|---|---|---|---|---|---|
| Parent Sales | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Parent Diversity | 0.000 (0.001) | 0.000 (0.001) | 0.001 (0.001) | 0.001 (0.001) | -0.014*** (0.001) | -0.011*** (0.001) |
| Industry Concentration | | | -0.023*** (0.001) | -0.023*** (0.001) | -0.023*** (0.001) | -0.022*** (0.001) |
| Industry Growth | | | 0.073*** (0.011) | 0.194*** (0.026) | 0.062*** (0.014) | 0.056* (0.031) |
| Industry Profitability | | | 0.222** (0.089) | -0.071** (0.033) | 0.232** (0.109) | -0.040 (0.032) |
| Relatedness | | | | | 0.152*** (0.002) | 0.125*** (0.002) |
| Share Strength | | | | | 0.071*** (0.017) | 0.057*** (0.015) |
| Intra Strength | | | | | 0.935*** (0.072) | 0.891*** (0.073) |
| Constant | -1.374*** (0.018) | -1.380*** (0.019) | -0.737*** (0.030) | -0.764*** (0.030) | -2.328*** (0.046) | -2.096*** (0.044) |
| -2LL | 31672.29 | 29175.30 | 30527.21 | 28143.27 | 22244.98 | 21753.25 |
| Model $\chi^2$ | 0.52 | 0.037 | 1145.60*** | 1032.07*** | 9427.83 | 7422.01 |
| Nagelkerke $R^2$ | 0.000 | 0.000 | 0.065 | 0.055 | 0.408 | 0.356 |

Logistic regressions of the probability of entry into a potential target industry by a diversifying firm. Standard errors in parantheses. ***, **, and * represent statistical significance at the 1, 5, and 10 percent levels, respectively. N=31357 for models marked 1981-83, N= 29037 for models marked 1983-85.

**Table 3: Change in the Probability of Entry for Given Changes of Independent Variables**

|  | + 1SD 1981-83 | + 1SD 1983-85 | + 10% 1981-83 | + 10% 1983-85 | + 100% 1981-83 | + 100% 1983-85 |
|---|---|---|---|---|---|---|
| Industry Concentr. | -5,46% | -5,40% | -0,93% | -0,91% | -7,25% | 7,19% |
| Industry Growth | 0,85% | 0,45% | 0,01% | 0,02% | 0,16% | 0,23 |
| Industry Prof | 0,55% | NA | 0,04% | NA | 0,46% | NA |
| Related-ness | 28,89% | 24,95% | 1,79% | 1,59% | 25,40% | 21,78% |
| Share Strength | 0,87% | 0,81% | 0,00% | 0,00% | 0,03% | 0,04% |
| Intra Strength | 3,24% | 3,13% | 0,19% | 0,17% | 1,99% | 1,81% |

The table shows the change in the probability of entry for different positive changes in the independent variables.   1 standard deviation change from the mean, 10% increase from the mean, and a 100% increase from the mean. The effects are calculated holding all other variables than the focal at their mean values. Results for both the 1981-83 sample and the 1983-85 sample are included.

**Table 4: Effects of adding each independent variable to Model 5**

| Variable | Change in -2 * Log Likelihood | Significance of change | Percent of total |
|---|---|---|---|
| Parent Sales | 0.411 | 0.521 | 0.00% |
| Parent Diversity | 157.74 | 0.000 | 1.73% |
| Industry Concentration | 702.45 | 0.000 | 7.70% |
| Industry Growth | 4.71 | 0.030 | 0.05% |
| Industry Profitability | 20.34 | 0.000 | 0.22% |
| Relatedness | 8055.07 | 0.000 | 88.32% |
| Share Strength | 17.55 | 0.000 | 0.19% |
| Intra Strength | 161.96 | 0.000 | 1.78% |

**Table 4 (Cont.): Effects of adding each independent variable to Model 6**

| Variable | Change in -2 * Log Likelihood | Significance of change | Percent of total |
|---|---|---|---|
| Parent Sales | 2.63 | 0.105 | 0.04% |
| Parent Diversity | 161.33 | 0.000 | 2.25% |
| Industry Concentration | 624.73 | 0.000 | 8.71% |
| Industry Growth | 3.16 | 0.075 | 0.04% |
| Industry Profitability | 1.79 | 0.181 | 0.02% |
| Relatedness | 6223.11 | 0.000 | 86.73% |
| Share Strength | 14.28 | 0.000 | 1.99% |
| Intra Strength | 144.18 | 0.000 | 2.01% |