The Rational Tradeoff between Corporate Scope and Profit Margins: The Role of Capacity-Constrained Capabilities and Market Maturity

Daniel Levinthal* and Brian Wu**

July 2007

We wish to thank the Mack Center for Emerging Technologies for generously supporting this research. We thank Ron Adner, David Collis, Glenn MacDonald, Nicolaj Siggelkow, Harbir Singh, and Sid Winter for their careful reading of a prior draft and valuable comments, as well as seminar audiences at the Tuck School, Dartmouth College, the Atlanta Competitive Advantage Conference, and the Academy of Management. In addition, we are appreciative of the very thoughtful comments of two reviewers and the Associate Editor.

*Wharton School, University of Pennsylvania

**Ross School of Business, University of Michigan

The Rational Tradeoff between Corporate Scope and Profit Margins: The Role of Capacity-Constrained Capabilities and Market Maturity

Abstract

While the contemporary literature on diversification from a resource perspective builds upon Penrose's (1959) idea of excess firm capabilities, the focus has been on the fungibility of resources across domains. Making a clear analytical distinction between scale-free capabilities and those that are capacity-constrained and need to allocated to one use or another helps to shift the discourse back to Penrose's (1959) original interest in the stock of organizational capabilities. The existence of capacity-constrained capabilities implies that rational diversification decisions should be based upon the opportunity cost of their use in one domain or another. The core result is that the recognition that resources and capabilities must be allocated among alternative uses provides a rational explanation for the divergence between total profits and profit margins. Firms make rational decisions to increase total profit via diversification when the industries in which they are currently competing become relatively mature. Due to the spreading of capacity-constrained capabilities across more segments, we may observe that firms' rational diversification actions lead to total profit growth but lower average returns. The model provides an alternative explanation for empirical observations regarding the diversification discount. We suggest that the self-selection mechanism in the diversification process may be based upon superior capabilities in a low value (existing market) context.

1. Introduction

The resource-based view of the firm has long recognized that firms diversify in order to exploit firm-specific resources[1] for which factor markets are imperfect (Penrose 1959; Teece 1982). The diversification literature along the lines of the resource-based view has largely focused on the fungibility of resources across domains (cf., Montgomery and Wernerfelt 1988). The literature has highlighted the degree to which the value of resources may be diminished as resources are leveraged in settings more distant from the original context in which the resource (e.g., brand-name or technical capability) was developed. Implicitly, resources tend to be viewed as having a *scale-free* property in the sense that the value of resources are assumed to be not reduced as result of the sheer magnitude of firm operations over which they are applied. Largely lost in the contemporary development of resource perspectives on diversification are some of the important insights of Penrose's early work that stressed the role of a firm's stock of resources in affecting diversification decisions. Many of the resources that may underlie a firm's diversification efforts, such as an effective management team or product development expertise in a particular domain, have the feature that they are *capacity constrained* --- that is, only so many goods and services can be generated by their use. At any point in time, these resources must be allocated among alternative activities, and the use of these resources in one activity precludes their use in other settings.[2] As a result, there are opportunity costs of using these capacity-constrained capabilities in one product market or another.[3]

The distinction between scale-free capabilities and capacity-constrained capabilities can be usefully illustrated by an analogy with public goods. Scale-free resources resemble public goods within

[1] "Resources" and "capabilities" are used interchangeably in this paper.
[2] The stock of capacity-constrained capabilities is certainly not fixed over time. However, the stock of capabilities at any point in time is capacity constrained as it takes time to accumulate rent-generating resources through flows of investments and activities, given the incompleteness and imperfectness of strategic factor markets (Dierickx and Cool 1989).
[3] Part of the reason why capacity-constrained capabilities are under-studied in the resource view is that this perspective emphasizes the role of scarcity of resources *in the factor markets* in explaining inter-firm heterogeneity (Barney 1986), but tends to ignore the fact that resources may also be scarce *within firm boundaries*.

firm boundaries in that simultaneous consumption of resources by multiple users does not diminish the value of consumption for any individual user. However, public goods can also be capacity-constrained in the sense that the value of consuming public goods for an individual user diminishes as the number of users increases. As a result, the consumption of the public good takes on a rivalrous quality. Examples of congestible public goods include roads, electricity and telephone networks, and swimming pools. Accordingly, it is important to efficiently allocate public goods across a certain number of users. The most familiar example in the business setting is a firm-specific management team (Penrose 1959; Slater 1980). While a superior management team can improve the productivities across all segments, the team also has to allocate its limited time and attention (Rosen 1982).

The above distinction between scale-free capabilities and those capabilities that must be allocated among alternative uses is critical because we believe that it is this later class of capabilities that capture the essence of Penrose's arguments regarding diversification. If capabilities are all scale-free, as is often implicitly assumed in the literature, issues of opportunity costs and resource allocation are inconsequential, as scale-free capabilities can always be leveraged in other areas and hence will always have "excess" capacity. Therefore, it is capacity-constrained capabilities that determine when excess capacity arises and how resources should optimally be allocated based on the considerations of opportunity cost. In this sense, we are making an analytical return to the original sensibility of Penrose's capability-based perspective on diversification. At the same time, adding to Penrose's emphasis on the internal growth of resources as the source of excess capacity in the context of demand environment that is implicitly assumed to be static demand, we examine how the dynamics of the demand environment influence the allocation of a firm's resources, its diversification efforts, and measures of performance. Specifically, since the criteria of carrying out an activity (or using the resource to generate other resources) are based upon the opportunity cost of capabilities (Rubin 1973; Slater 1980), a complete account of excess capacity of capabilities should take into account not only internal growth in firm-specific capabilities but also the change in external opportunities across different markets. Underutilized capacity becomes available when the growth opportunities in the current market cannot keep pace with the internal growth of capabilities. The maturity of the current market relative to other potential markets[4] could either reduce the value of applying capacity-constrained capabilities in the current market or raise the opportunity cost of not applying some of these capabilities in related product markets. It is in this sense that resources become “underutilized” or “excess”.[5] Alternatively, if the current market continues to offer sufficiently favorable opportunities, it will not be economically rational to divert capacity-constrained capabilities into other industries as long as there is any imperfect fungibility in the value of capabilities when applied to other domains.

Building on these issues concerning firms’ internal resource base and their external product market environment, we develop a basic economic model that provides a rational explanation of firms’ diversification behavior in trading off profit margins for corporate growth. Largely ignored by the literature is the fact that rational diversification decisions imply that firms seek to increase total profit, but not necessarily their profit margin or market-to-book value --- with the later two measures being among the more common performance measures used in the diversification literature (Palich, Cardinal, and Miller 2000). Firms make rational decisions to increase total profit via diversification when the industries in which they are currently competing become relatively mature. In this process, however, firms need to allocate their capacity-constrained resources away from the current business to the new one. Due to the spreading of these capabilities across more segments, we may observe that firms' rational diversification actions lead to total profit growth but lower average returns. This result adds to early work applying the resource view of the firm to the question of diversification within the strategy literature (e.g., Montgomery and Wernerfelt 1988), which also suggests that diversification may not conflict with value maximization because firms diversify only when *marginal* returns are positive. Montgomery and

[4] The relative maturity of the current market could arise either from the decline of the current market or from the fast growth of other markets. An example of the former case is the defense industry after the mid 1980s (Anand and Singh 1997) while an example of the latter case is the mature desktop PC market in comparison with the rapidly growing hand-held device market.
[5] Penrose (1959) also recognized the role of the lumpiness of investment in leading to excess capacity. This type of excess firm capabilities differs from the capacity-constrained capabilities highlighted in this paper in that the reallocation of such excess, indivisible resources into a new market does not hurt firms’ efficiency or productivity in the existing markets and therefore involves zero opportunity cost.

Wernerfelt (1988) show that a wider level of diversification can lead to lower *average* rents (Tobin's *q*) due to the imperfect fungibility of firm-specific factors. We find that the decline in average returns may arise from the reallocation of capacity-constrained capabilities even in the absence of any imperfect fungibility of firm-specific capabilities as they are allocated to new product markets.

The remainder of the paper is organized as follows. We develop further the notion of capacity-constrained capabilities that must be allocated among alternative uses and its contrast with, what we term, scale-free capabilities in Section 2. Section 3 sets up the model by linking the theories regarding diversification, capabilities, and demand conditions. Section 4 develops the analysis regarding ex-ante diversification decisions and ex-post performance. Section 5 discusses the implications of the model and provides some broader conclusions of the work.

2. Capacity-Constrained Capabilities and Diversification

The following two-by-two table illustrates the relationship between the current study and the existing literature.[6] The existing literature focuses on scale-free capabilities, such as technical know-how and reputation, which lead to economies of scope or synergies in the diversification process because they "display some of the characteristics of a public good in that it may be used in many different non-competing applications without its value in any one application being substantially impaired (Teece 1980: p.226)". The recognition of scale-free capabilities has had profound influence on both academic research and industry practice, as it highlights the role of knowledge and competence as strategic assets (Winter 1987). Indeed, in their study of replication as strategy, Winter and Szulanski (2001) provide a paradigmatic example of a scale-free capability. They define the Arrow core as the informational endowment a firm extracts from an original setting which they can replicate to other settings. The distinctive property of such information-like resource is that "unlike any resource that is rivalrous in use, an information-like resource is infinitely leverageable… it does not have to be withdrawn from one use to be applied to another (Winter and Szulanski 2001: p.741)."

[6] We would like to thank David Collis for suggesting this table.

	Capacity-constrained	Scale-free
High fungibility	E.g., team of auditors; power generation equipment	E.g., brand-name; computer operating system
Low fungibility	E.g., personnel with specific technical expertise; steel plant	E.g., patent; customer relationship

Table 1: Dimensions of capabilities

One important property of scale-free capabilities is their fungibilitiy, which is probably one of the most studied questions in corporate strategy. Rumelt (1974), in a pioneering examination of this issue, showed that firms pursuing related diversifications outperform those pursuing unrelated one. This basic finding has been reconsidered with a variety of different measures of relatedness, but the general results have stood-up (Montgomery and Wernerfelt 1988; Markides and Williamson 1994; Robins and Wiersema 1995).

While Penrose's emphasis on a firm's stock of capabilities as a basis for diversification has played a secondary role to the consideration of the fungability of resources, an awareness that there may be capacity constraints on resources has not been absence in the literature. In his brief discussion on the limits to diversification economies, Teece (1982) suggests that "knowhow is generally not embodied in blueprints alone; the human factor is critically important in technology transfer. Accordingly, as the demands for sharing knowhow increase, bottlenecks in the form of over-extended scientists, engineers, and managers can be anticipated (Teece 1982: 53)." Recent empirical work in both finance and management has also provided evidence that is consistent with the existence of capacity-constrained capabilities. Using plant-level observations from the Census data, Schoar (2002) observes that after a firm diversifies into a new industry by acquiring a plant, the incumbent plants will incur decrease in productivity, while the acquired plants increase productivity after being acquired. Roberts and McEvily (2005) show that entering a new pharmaceutical product market reduces a firm's performance in the current markets.

Consider the following example of these arguments. As the strongest player in the microprocessor industry, Intel has been experiencing sluggish growth in the PC microprocessor market due to saturated demand and increasing competition from Advanced Micro Devices. In order to spur growth, Intel has sought to extend its reach beyond the PC microprocessor industry into the cell phone and communications industry. The maturity of the PC microprocessor market has made the opportunities of using its capabilities in other industries more attractive and, correspondingly, the opportunity cost of staying focused has risen. At the same time, diversification requires Intel to allocate its scarce resources into these new segments. Consequently, our theory would predict that, as a whole, Intel's sales and total profit will grow, but its average return will decline, reflecting both the shifting away of firm-specific resources from the development and manufacturing of microprocessors for PCs and the possible reduced efficacy of these same resources in the related product markets into which the firm is diversifying.[7]

Based upon the above reasoning, we develop the following arguments regarding diversification efforts. First, it is important to distinguish between diversification efforts based on capacity-constrained resources versus scale-free resources. A scale-free resource, such as brand-name, faces limits on the breadth of its fungibility (for instance, how broadly fungible is a given brand-name) but not on its extent of application (for instance, the number of markets in which a given brand can be applied assuming fungibility is perfect). In contrast, the application of capacity-constrained capabilities is driven by a logic of opportunity costs --- on the margin is the greatest value of this firm-specific capability realized within the current product market context or in diversifying to a new context. This opportunity cost is, in turn, importantly affected by the size, growth, and competitive conditions in alternative product markets.

[7] Of course, another basis for the decline in average return is the shift from a market in which Intel has a dominate position to markets that may be more competitive; however, the fact that Intel is entering these more competitive, but more rapidly growing markets, is further testimony to the need to reallocate capacity constrained capabilities.

Furthermore, there is the question of a firm's viability in a given product market. Viability addresses the question of whether sufficient capabilities are allocated to a given market so as to create a positive net value added, defined as the difference in consumers' willingness to pay over the existing competitors' product offerings minus the firm's costs of operating in this market. Thus, when there are multiple segments, it can be shown that the range of diversification activity is constrained by the total stock of capabilities. This result supplements the insight in the strategy literature that the imperfect fungibility of scale-free capabilities restricts corporate scope (e.g., Montgomery and Wernerfelt 1988).

This analysis also provides new insights for explaining the observed cross-sectional diversification discount (Lang and Stulz 1994; Berg and Ofek 1995). Agency theorists suggest that diversification destroys value for reasons such as managers' empire building behavior that aims to increases their own status, power, and pecuniary compensation (e.g., Jensen 1986). Recently, however, there has been a growing literature in the corporate finance field suggesting that a diversification discount arises even when firms are value maximizers. Econometrically sophisticated analyses of the profitability of diversified firms (e.g., Campa and Kedia 2002; Villalonga 2004) indicate that there is something systematically different about firms that diversify. It is this endogenous selection into the act of diversification, rather than diversification per se, that leads to diversification discount:

> "… the failure to control for firm characteristics that lead firms to diversify and be discounted may wrongly attribute the discount to diversification instead of the underlying characteristics. For example, consider a firm facing technological change, which adversely affects its competitive advantage in its industry. This poorly performing firm will trade at a discount relative to other firms in the industry. Such a firm will also have lower opportunity costs of assigning its scarce resources in other industries, and this might lead it to diversify. If poorly performing firms tend to diversify, then not taking into account past performance and its effect on the decision to diversify will result in attributing the discount to diversification activity, rather than to the poor performance of the firm." (Campa and Kedia 2002: p. 1732)

In existing analytical explanations of this empirical finding that, controlling for endogeniety in diversification behavior there is no diversification discount (e.g., Gomes and Livdan 2004), the act of diversification is interpreted as a "signal" that the firm has relatively few ex-ante capabilities and is diversifying due to the correspondingly low rates of return in its initial markets. In the presence of diminishing returns to production, firms with lower productivity will reach their optimal size in the incumbent segment at a lower size level than those firms with higher productivity and, as a result, firms with lower productivity are more likely to diversify.

We agree with these recent empirical findings that there is something systematic about those firms that "sort" themselves into a positive diversification decision. However, the above analytical explanation is not fully consistent with the well-evidenced proposition in the strategy field that firms with more relevant capabilities (R&D capabilities or marketing capabilities) tend to enter a new field earlier and perform better (e.g., Klepper and Simons 2000; Mitchell 1989).[8] In contrast, in the spirit of the long-standing treatment of diversification in the strategy literature, we suggest that the "something different" is not that these firms are a "bad type" and are lacking in capabilities. Rather, these are firms with relatively superior capabilities and the bad "signal" may be a statement about the market contexts in which these firms are operating, such as demand maturity, rather that a statement about the firm's relative lack of capabilities. In the pursuit of the best use of the firm's capacity-constrained resources, there is some allocation of resources away from established markets and, at the sacrifice of profit margins but not total profits, a shift of these resources to new markets. Thus, both our model and those developed in the corporate finance literature are consistent with the empirical finding regarding the diversification "discount"; however, the two explanations differ in their predictions as to which firms (more or less capable) are likely to be more or less diversified.

3. Model Structure[9]

In our initial analysis, we model a firm's diversification decision with regard to two market segments indexed by *m* (The initial segment $m = I$ and the new segment $m = N$); this is expanded to a set of *M* possible markets in the subsequent section. Production in each segment is described as[10]

[8] See Helfat (2003) for a review of the capability leverage literature.

[9] We would like to thank Glenn MacDonald for his input on the model structure and analysis.

[10] We assume that the production function exhibits constant returns to production scale. This assumption of constant return to scale allows us to demonstrate that diversification can arise from the change in demand conditions in the absence of diminishing return to scale, in contrast to recent work in corporate finance (cf., Gomes and Livdan, 2004 and Maksimovic and Phillips, 2002) in which diminishing returns acts as an underlying driver for diversification.

$$Q_m = \gamma_m t_m T k_m \tag{1}$$

where γ_m is the firm's scale-free capabilities, t_m is the share of the firm's total capacity-constrained capabilities, T, that must be divided among activities, and capital input k_m is the sole purchased input, whose unit price is r. The amount of k_m needed to produce Q_m, given t_m, is therefore $\frac{Q_m}{\gamma_m t_m T}$, with total cost $\frac{rQ_m}{\gamma_m t_m T} = mc_m \times Q_m$, where

$$mc_m \equiv \frac{r}{\gamma_m t_m T} \tag{2}$$

Note that (i) scale-free capabilities, γ_m, and capacity-constrained capabilities, T, are firm specific and subject to imperfect input markets (Teece 1982); (ii) providing more of the capacity-constrained capability reduces both total and marginal cost since it substitutes for the purchased input; (iii) mc_m is decreasing in t_m and increasing to infinity as t_m approaches zero; (iv) when the firm enters a new market, it needs to invest in new capital.

The demand side of segment m consists of s_m consumers, all having the same willingness to pay, w_m. Each market is a Bertrand duopoly, where the competing firm has constant marginal cost c_m. So whenever the firm allocates the capacity-constrained capabilities such that $mc_m < c_m$, the firm serves the whole market, charges a price of $p_m = c_m$, and has profits $s_m(p_m - mc_m)$.

Should the firm engage in both activities, its profit is:

$$s_I(p_I - mc_I) + s_N(p_N - mc_N)$$

or

We have shown that the results hold with a more general production function which allows for diminishing return on k_i. These results can be made available upon request; however, the assumption of constant returns to scale provides for a more intuitive statement of the results.

$$s_I(p_I - \frac{r}{\gamma_I t_I T}) + s_N(p_N - \frac{r}{\gamma_N t_N T}) \quad (3)$$

where $\gamma_N = (1-\delta)\gamma_I$, which indicates that scale-free capabilities, γ_I, is not perfectly fungible, or that the effectiveness of scale-free capabilities, γ_I, diminishes by a factor δ ($0 \le \delta < 1$) when γ_I is applied to the new segment *N*.[11]

It should be noted that the profit expressed above characterizes the revenue in excess of the amount required to pay for the capital investment needed to generate the profit. Therefore, it captures the notion of return on invested capital and should be interpreted as scarcity rent arising from owning scarce firm-specific factors (Lindenberg and Ross 1981; Winter 1995). Observing that when we write profits as revenue less cost, profit function (3) can be transformed as

$$(p_I s_I + p_N s_N) - r(\frac{s_I}{\gamma_I t_I T} + \frac{s_N}{\gamma_N t_N T}) \quad (4)$$

from which it is clear that choosing t_I and t_N to maximize profit is the same as choosing these values to minimize costs.

As a result, assuming the optimal t_I and t_N are both strictly positive, the firm's problem is:

$$\min\{\frac{s_I}{\gamma_I t_I T} + \frac{s_N}{\gamma_N t_N T} \mid t_I + t_N = 1, \frac{r}{\gamma_N t_N T} \le p_N, \text{and } \frac{r}{\gamma_I t_I T} \le p_I\} \quad (5)$$

Note that it may not be possible for the firm to engage in both activities simultaneously, let alone optimal. Moreover, for finite p_I and p_N, the optimal allocations of the capacity-constrained capabilities t_I^* and t_N^*, are bounded away from zero, i.e., to get into the new activity will require a discrete reduction in the capacity-constrained capabilities employed in the initial activity.

[11] As we point out in Table 1, capacity-constrained capabilities are also subject to the issue of imperfect fungibility, but the distinctive feature of capacity-constrained capabilities is that they must be allocated across alternative uses. In order to highlight the issue of allocation, we assume that capacity-constrained capabilities are perfectly fungible; however, so that the model more closely corresponds to Montgomery and Wernerfelt (1988), we allow for the scale-free capability to have imperfect fungibility across alternative uses.

4. Analysis

In this section, we first analyze the allocation problem of capacity-constrained capabilities across product markets based on the consideration of opportunity costs. We then use the result of this allocation problem to identify the demand thresholds for firms to stay focused or become diversified and determine the extent of firm diversification. We then shift the analysis from the choice of product markets in which to enter to a consideration of diversification performance and demonstrate that a profit maximizing firm may diversify in ways that reduce the firm's profit margin. In doing so, we provide a profit-maximizing explanation for the well-documented diversification discount. Throughout the analysis, we emphasize the role of capacity-constrained capabilities and its distinction from scale-free capabilities.

4.1. The allocation of capacity-constrained capabilities

Figure 1 illustrates the firm's optimization problem of minimizing costs of production across the two markets, subject to the constraint of the total capabilities available to the firm.

Insert Figure 1 about here

First, we solve the firm's minimization problem (5). The (t_I, t_N) pairs satisfying the constraints are along the straight line $t_I + t_N = 1$. Level sets of the objective function are iso-cost curves with cost falling when moving outward. The slope of the iso-cost curve is

$$\frac{dt_N}{dt_I} = -\frac{\gamma_N}{\gamma_I}\frac{s_I}{s_N}\left(\frac{t_N}{t_I}\right)^2$$

As a result, the optimal division of the capacity-constrained capabilities can be characterized by point A. At this point, the slope of the iso-cost curve is equal to that of the constraint line, i.e.,

$$\frac{dt_N}{dt_I} = -\frac{\gamma_N}{\gamma_I}\frac{s_I}{s_N}\left(\frac{t_N}{t_I}\right)^2 = -1$$

The optimal allocation of capacity-constrained capabilities can be readily solved as[12]

$$t_I^* = \frac{\sqrt{(1-\delta)s_I}}{\sqrt{(1-\delta)s_I} + \sqrt{s_N}} \text{ and } t_N^* = \frac{\sqrt{s_N}}{\sqrt{(1-\delta)s_I} + \sqrt{s_N}} \qquad (6)$$

When there are no interior solutions, either $t_I^* = 1, t_N^* = 0$ or $t_I^* = 0, t_N^* = 1$.

One straight-forward property of the equilibrium allocation of capabilities is that $\frac{\partial t_N^*}{\partial s_I} < 0$, $\frac{\partial t_N^*}{\partial s_N} > 0$, and $\frac{\partial t_N^*}{\partial \delta} > 0$ when there are interior solutions.

Proposition 1: If it is optimal for firms to allocate positive amounts of capacity-constrained capabilities across two segments, the fraction of allocation is determined by the relative size of the two market segments and fungibility of scale-free capabilities. The fraction allocated to the new segment is (i) decreasing with the size of the initial segment, (ii) increasing with the size of the new segment, and (iii) increasing with the degree to which the effectiveness of scale-free capabilities diminishes in the new segment.

Proposition 1(i) and 1(ii) demonstrate that the allocation of internal capabilities is determined by the demand conditions, or opportunity cost. Firms tend to allocate more scarce capabilities towards more favorable markets. Proposition 1(iii) demonstrates the relationship between fungibility and resource allocation. This property is somewhat counterintuitive in that firms allocate more scarce capabilities toward a new market when the effectiveness of scale-free capabilities diminishes to a larger degree. This result stems from the fact that the imperfect fungibility of scale-free capabilities has two effects. On the one hand, as we will show in the next subsection, a larger degree of imperfect fungibility decreases the attractiveness of the new segment as a diversification target. In other words, it requires a larger market size for a market with a larger degree of imperfect fungibility to become attractive. Therefore, firms are less likely, other things being equal, to enter a more distant segment, consistent with prior arguments in the strategy literature. On the other hand, once the new segment becomes sufficiently attractive so as to

[12] It should be noted that while competitors' cost efficiency does not affect the continuous allocation of capacity-constrained capabilities to a certain segment, it does influence the boundary conditions that determine whether a firm is viable to enter this segment.

make entry attractive, conditional upon entry, a larger degree of imperfect fungibility requires a greater allocation of capacity-constrained capabilities to compensate for this diminished effectiveness in capabilities.

4.2. Capacity-constrained capabilities and the extent of diversification.

We first specify the viability condition for a firm to operate in a given market, which will be used repeatedly in the following analysis.

Lemma 1:A firm is viable in a market m if and only if its profit margins $p_m - \frac{r}{\gamma_m t_m^* T} > 0$.

Lemma 1 implies that a firm is viable in a given market only when the optimal allocation of capabilities makes its profit margins positive, where this optimal allocation, t_m^*, is based upon the considerations of opportunity costs which are in turn determined by relative market size as specified in equation (6) above. For instance, point *B* in Figure 1 is an optimal solution based on equation (6), but it is not viable in the new market, and therefore the firm will still focus on the initial market. This is the same for point *C*.

In order to demonstrate the role of demand conditions in determining firms' diversification decisions, we first identify the demand threshold for a given firm to enter into the new segment *N*, $\bar{s}$, and the threshold for the firm to exit the initial segment *I*, $\underline{s}$. We express the threshold values as the ratio of the size of the initial market over that of the new market, or $\frac{s_I}{s_N}$.

According to Lemma 1, to derive the threshold for the firm to enter into the new segment *N*, $\bar{s}$, we can let

$$p_N - \frac{r}{\gamma_N t_N^* T} = p_N - \frac{r}{\gamma_N \frac{T\sqrt{s_N}}{\sqrt{(1-\delta)s_I} + \sqrt{s_N}}} = p_N - \frac{r}{(1-\delta)\gamma_I \frac{T\sqrt{s_N}}{\sqrt{(1-\delta)s_I} + \sqrt{s_N}}} = 0$$

We can then solve for

$$\bar{s} = \frac{s_I}{s_N} = \frac{1}{(1-\delta)}[\frac{1}{r} p_N (1-\delta)\gamma_I T - 1]^2 \tag{7}$$

It is then straightforward to show the following relationships: $\frac{\partial \bar{s}}{\partial \gamma_I} > 0$, $\frac{\partial \bar{s}}{\partial T} > 0$, $\frac{\partial \bar{s}}{\partial \delta} < 0$, and

$\frac{\partial \bar{s}}{\partial p_N} > 0$.

Proposition 2: There exists a critical threshold of the market size of the initial segment relative to the new segment, defined by equation (7). Firms stay focused when the initial segment's relative size remains above this threshold and becomes diversified when the initial segment's relative size drops below the threshold. For a given firm, the initial segment reaches this diversification threshold at a higher relative market size when (i) it has more scale-free capabilities and capacity-constrained capabilities; (ii) the degree to which the effectiveness of scale-free capabilities diminishes in the new segment is lower; and (iii) competitors' cost efficiency in the new segment is lower.

Similarly, we can derive the threshold $\underline{s}$ for the firm to finally exit the initial segment *I*.

$$\underline{s} = \frac{s_I}{s_N} = \frac{1}{(1-\delta)} \frac{1}{[\frac{1}{r} p_I \gamma_I T - 1]^2} \tag{8}$$

Again, it is straightforward to show the follow relationships: $\frac{\partial \underline{s}}{\partial \gamma_I} < 0$, $\frac{\partial \underline{s}}{\partial T} < 0$, $\frac{\partial \underline{s}}{\partial \delta} > 0$, and

$\frac{\partial \underline{s}}{\partial p_I} < 0$.

Proposition 3: There exists a critical threshold of the market size of the initial segment relative to the new segment, defined by equation (8).Firms stay diversified when the initial segment's relative size remains above the threshold and exits the initial segment and focuses on the new segment when the initial segment's relative size drops below this threshold. For a given firm, the initial segment reaches this exit threshold at a lower relative market size when (i) it has more scale-free capabilities and capacity-constrained capabilities; (ii) the degree to which the effectiveness of scale-free capabilities diminishes in the new segment is lower; and (iii) competitors' cost efficiency in the initial segment is lower.

Proposition 2 and 3 can be summarized by Figure 2.

Figure 2: Diversification dynamics associated with market maturity in the initial segment

When the size of the initial segment is large relative to the new segment, firms stay focused in the initial segment. As the demand in the initial segment decreases relative to the new segment, firms will first diversify into the new segment while staying in the initial one. As the relationship in relative market size becomes more extreme, firms will exit the initial segment and focus exclusively on the new segment. This finding is similar to the argument of Helfat and Eisenhardt (2004) who suggest that firms may exhibit inter-temporal economies of scope by redeploy resources across markets that experience rapid changes in technology and demand.

Proposition 2 and 3 emphasizes the role of *relative* demand in the initial market over the new market, as opposed to the *absolute* demand in the new market, in determining firms' diversification decisions and therefore provide a test to disentangle capacity-constrained capabilities that must be allocated among alternative uses from scale-free capabilities. When capabilities must be allocated among alternative uses, the relative maturity in the initial market raises the opportunity cost of exclusively focusing these capabilities in the current market and in turn, from an optimal allocation perspective, frees up some of these capabilities for their use in a new market. Therefore, firms' decisions to enter the new market are affected by the relative demand conditions across markets. In contrast, assuming capabilities are all scale-free, firms only need to concern themselves with whether the new market is sufficiently attractive and their decisions to enter the *new* market are *not* affected by the demand conditions in the *initial* market.

Proposition 2 and 3 also demonstrate that more capable firms diversify in that the range of relative market size for the firm to become diversified increases with the magnitude of the firm's capabilities (Propositions 2 (i) and 3 (i)).

The critical role of capacity-constrained resources that must be allocated across alternative uses can be further highlighted by the extreme case that firms that lack sufficient capabilities will never become diversified. Note that Figure 2 implicitly assumes that $\bar{s} > \underline{s}$. This assumption holds only when a firm possesses a sufficient amount of capacity-constrained capabilities. In other words, a minimum amount of capacity-constrained capabilities are necessary to establish competitive viability in both product markets simultaneously. Let $\bar{s} = \underline{s}$, we can solve for this minimum amount of capabilities as

$$\bar{T} = \frac{r}{\gamma}\left(\frac{1}{p_I} + \frac{1}{(1-\delta)\,p_N}\right) \tag{9}$$

When $T < \bar{T}$, the assumption that $\underline{s} < \bar{s}$ does not hold as the firm lacks sufficient capacity-constrained capabilities to become a diversified firm. Therefore, as the relative size of the two markets changes, firms with sufficient capabilities will go through a process of diversification, while those without sufficient capabilities will either focus on the initial segment or switch all resources from one segment to the other, independent of demand conditions.

The role of capacity-constrained capabilities that must be allocated across uses in establishing competitive viability can be illustrated by an adaptation of Figure 1 (Figure 3). When T equals the threshold $\bar{T}$ defined in equation (9), the pair of straight lines representing the viability conditions (defined by Lemma 1) in each market intersect with the straight line representing the capability constraint at the same point *D*. When $T > \bar{T}$ (e.g., the pair of dotted lines with $T = T_1$ in Figure 3), the viability condition lines move toward zero. As a result, firms diversify when the relative demand conditions are such that the optimal allocation (t_I^*, t_N^*) falls between D_1 and D_2 along the straight line $t_I + t_N = 1$. In this case, t_I^* and t_N^* not only are both strictly positive but also satisfy viability conditions in both markets. Moreover, as T further increases, the range of relative market size for diversification, which corresponds to the distance between by D_1 and D_2, further increases. This also illustrates the results in Propositions 2 (i) and 3 (i). In contrast, when $T < \bar{T}$ (e.g., the pair of dotted lines with $T = T_2$ in Figure 3, the viability condition lines move further outward. In this case, firms are never able to diversify because there does not exist any point along the capability constraint line that can simultaneously satisfy the viability conditions in both markets.[13]

Insert Figure 3 about here

The above logic can be extended to the case in which there are multiple potential markets into which firms can diversify and the range of diversification activity is endogenously determined by the firms' total capabilities. This analysis is useful because it conceptually distinguishes our model from other arguments put forward to account for the presence of a diversification discount in the context of firm profit maximizing behavior. Montgomery and Wernerfelt (1988) suggest that the number of segments is determined by the fungibility of capabilities. The effectiveness of capabilities diminishes when they are leveraged in settings more distant from the original context in which they were developed. Therefore, a firm will continue diversifying into the next best new opportunity until the point when the marginal rent becomes zero. Our analysis extends this insight and incorporates the role of capacity-constrained capabilities and the need to establish competitive viability in individual product markets. As a result, even holding constant the degree to which capabilities are diminished in effectiveness when they are applied outside the initial market domain, the range of diversification is influenced by the magnitude of the firm's capabilities. From this perspective, our model further highlights the difference between capacity-constrained and scale-free capabilities.

The above distinction can be illustrated with the following experiment. Consider an analysis in which there are *M* product markets each of which is equally "distant" from the initial product market. If diversification were driven solely by the fungibility of capabilities, a firm would diversity into all *M* markets;[14] however, if capabilities are capacity-constrained and need to be allocated across markets and,

[13] In the extreme case, when T becomes even smaller, the firm will not be viable in either or both markets even when all capabilities are focused in one market.

[14] The model structure developed by Maksimovic and Phillips (2002) also implies that, driven by the assumption of diminishing returns to production, firms equally diversify into all segments. Gomes and Livdan (2004) modify this

further, that firms need to establish competitive viability in each of these *M* markets, then a firm will only diversify into a subset of them. The above logic is formalized in the following case of *M* heterogeneous markets. The necessary capability threshold for these *M* distinct segments to be viable simultaneously is (see Appendix 1 for a proof):

$$\overline{T} = \frac{r}{\gamma}\sum_{m=1}^{M}\frac{1}{p_m(1-\delta_m)} \quad (10)$$

Combining the above analysis, we have Proposition 4:

Proposition 4: The amount of capacity-constrained capabilities limits the maximum number of distinct segments that can possibly be supported. The capability threshold for a given number of markets to be viable is defined by equation (10).

Proposition 4 establishes the relationship between firm capabilities, the fungibility of capabilities, and the number of segments in the diversification portfolio. The driver of such relationship is the need to allocate capacity-constrained capabilities. The viability conditions are determined by consumers' willingness to pay and competitors' cost efficiency, which in turn determines the minimal level of capacity-constrained capabilities that must be applied to a given market. Moreover, apart from the viability conditions, it is market size that determines the opportunity cost associated with the use of capacity-constrained capabilities and in turn determines which market segment to diversify into and how much capacity to allocate to that market.

4.3. The rational tradeoff between total profit and profit margin

This subsection examines the effect of diversification on the rational tradeoff between total profit and profit margin. We demonstrate that when firms seek to *rationally* maximize total profit by diversifying into a new segment, their capacity-constrained capabilities will be spread across segments and therefore their average profit margin will inevitably decline, controlling for market conditions of all

---

structure and introduce the presence of a fixed cost associated with each market entry so as to generate an upper bound on the number of product markets in which a firm participates.

segments. Since profit margin, or market-to-book value, is among the more common performance measures (far more common than total profit) in empirical research, observations of declining average return might suggest an agency argument that executives are expanding their firms at the expense of shareholders. Our model suggests that the tradeoff between total profit and profit margin does not necessarily result from an agency problem between management and shareholders. Rather, the decline in profit margins and the well documented diversification discount may well be consistent with value maximization. A given firm diversifies into the new segment when demand conditions are such that the minimization problem (5) has an interior solution. We denote this interior solution as the diversification strategy $(0 < t_{iI}^* < 1, 0 < t_{iN}^* < 1, t_{iI}^* + t_{iN}^* = 1)$. The total profit associated with the diversification strategy is

$$\pi = s_I (p_I - \frac{r}{\gamma_I t_I^* T}) + s_N (p_N - \frac{r}{\gamma_N t_N^* T}) \tag{11}$$

Since this diversification strategy $(0 < t_I^* < 1, 0 < t_N^* < 1, t_I^* + t_N^* = 1)$ is the optimal solution to the minimization problem (5), obviously it is associated with a larger profit level than any other solutions, including the strategy to stay focused either in the initial segment $(t_I = 1, t_N = 0)$ or in the new segment $(t_I = 0, t_N = 1)$. Therefore, under certain demand conditions, it is better off in terms of total profit maximization for a firm to diversify.

However, under the same demand conditions, the diversification strategy is associated with a lower profit margin than the weighted average of the focus strategy in the initial segment and that in the new segment, which is a standard benchmark used to compare the performance of diversified and focused firms (Lang and Stulz 1994; Berg and Ofek 1995).[15]

[15] In practice, the diversification discount is calculated by comparing a diversified firm's value to its imputed value, where the latter is the weighted sum of the median value of single-segment firms of similar size in each of the segments in which it operates (Lang and Sultz 1994). This method has been criticized for ignoring the systematic difference between diversified firms and focused firms. However, by using the same firm with the same stock of capabilities and market conditions, the imputed value calculated below in Equation (12) and (13) does not suffer from this criticism.

The profit margin associated with the diversification strategy is

$$\frac{\pi}{p_I s_I + p_N s_N} = \frac{s_I}{p_I s_I + p_N s_N}(p_I - \frac{r}{\gamma_I T t_I^*}) + \frac{s_N}{p_I s_I + p_N s_N}(p_N - \frac{r}{(1-\delta)\gamma_I T t_N^*}) \quad (12)$$

When firms focus on the initial segment, the profit margin is $\frac{s_I(p_I - \frac{r}{\gamma_I T})}{p_I s_I}$. When firms focus on the new segment, the profit margin is $\frac{s_N(p_N - \frac{r}{(1-\delta)\gamma_I T})}{p_N s_N}$. Therefore, the weighted average of the profit margin of the two focus strategies, with relative sales $\frac{p_I s_I}{p_I s_I + p_N s_N}$ and $\frac{p_N s_N}{p_I s_I + p_N s_N}$ as the respective weights, is

$$\frac{p_I s_I}{p_I s_I + p_N s_N}\frac{s_I(p_I - \frac{r}{\gamma_I T})}{p_I s_I} + \frac{p_N s_N}{p_I s_I + p_N s_N}\frac{s_N(p_N - \frac{r}{(1-\delta)\gamma_I T})}{p_N s_N}$$

$$= \frac{s_I}{p_I s_I + p_N s_N}(p_I - \frac{r}{\gamma_I T}) + \frac{s_N}{p_I s_I + p_N s_N}(p_N - \frac{r}{(1-\delta)\gamma_I T_i}) \quad (13)$$

Comparing Equation (12) with Equation (13), we see that $(p_I - \frac{r}{\gamma_I T t_I^*}) < (p_I - \frac{r}{\gamma_I T})$ and $(p_N - \frac{r}{(1-\delta)\gamma_I T t_N^*}) < (p_N - \frac{r}{(1-\delta)\gamma_I T})$, since the optimal diversification strategy is $0 < t_I^* < 1, 0 < t_N^* < 1$. As a result, the profit margin of the diversification strategy defined in Equation (12) is smaller than the profit margin of the focus strategy defined in equation (13).

We further decompose the profit margin of the diversification strategy into two parts---the weighted average of the two focus strategies and the discount due to the spreading of capacity-constrained capabilities. This decomposition allows us to identify the sources of the declining profit margin associated with diversification.

According to Equation (6), $t_I^* = \frac{\sqrt{(1-\delta)s_I}}{\sqrt{(1-\delta)s_I} + \sqrt{s_N}}$ and $t_N^* = \frac{\sqrt{s_N}}{\sqrt{(1-\delta)s_I} + \sqrt{s_N}}$. Therefore, equation (12), the profit margin for the diversification strategy, can be transformed as

$$\frac{\pi}{p_I s_I + p_N s_N} = [\frac{s_I}{p_I s_I + p_N s_N}(p_I - \frac{r}{\gamma_I T}) + \frac{s_N}{p_I s_I + p_N s_N}(p_N - \frac{r}{(1-\delta)\gamma_I T})] - 2\frac{\sqrt{s_I s_N}}{p_I s_I + p_N s_N}\frac{r}{\sqrt{(1-\delta)}\gamma_I T} \quad (14)$$

Notice that the first term (in the square brackets) in the above equation is exactly the weighted average of two focus strategies in Equation (13), an inspection of which suggests that, in the absence of the capability constraints, the profit margin of the diversification strategy declines with δ, the degree to which the effectiveness of scale-free capabilities diminishes in the new segment. This is the case studied in Montgomery and Wernerfelt (1988) who account for a decline in profit margin or market to book value as firms apply, what we term, scale-free capabilities in increasingly distant markets. Adding to this consideration of imperfect fungibility, the second term captures the discount due to the spreading of capacity-constrained capabilities. Therefore, Equation (14) provides a more complete picture of the diversification discount from the resource-based perspective by incorporating the effect of capacity-constrained resources that need to be allocated across applications.

Combining the above analyses regarding total profit and profit margin, we have the following proposition.

Proposition 5: Profit-maximizing diversification strategy is associated with a larger total profit than focus strategies, but a lower profit margin than the weighted average of focus strategies.

Proposition 5 suggests that the existence of a diversification discount does not necessarily result from agency behavior that deviates from profit maximization. The spreading of capacity-constrained capabilities implies that the profit margin will be “sacrificed” to some extent in the pursuit of total profit maximization.

Finally, we link the profitability analyzed in the current analysis to the Tobin's *q* widely used in the empirical analysis. In this stylized one-period model,[16] market value can be represented by total revenue $p_m s_m$ and total capital is the total units of capital times its capital price $r\frac{s_m}{\gamma_m T}$. Therefore, Tobin's *q* can be represented as

$$q = \frac{p_m s_m}{\frac{r}{\gamma_m T} s_m} = \frac{p_m}{\frac{r}{\gamma_m T}} = \frac{p_m}{mc_m} \tag{15}$$

Note that $\frac{r}{\gamma_m T}$ should be viewed as average cost (which equals marginal cost in this special case, according to Equation (2)). In this sense, Tobin's *q* capitalizes the difference between price and average cost (Lindenberg and Ross 1981; Winter 1995).[17] In fact, if we base our analysis upon Tobin's *q* as defined above, the discount arising from the spreading of capacity-constrained capabilities can be expressed as[18] (see Appendix 2 for a proof)

$$\frac{2\sqrt{\frac{s_I s_N}{1-\delta}}}{s_I + \frac{s_N}{1-\delta}} \tag{16}$$

Therefore, whether the average return to capital is measured as profit margin (Equation (14)) or by Tobin's *q* (Equation (16)), we see that profit maximizing diversification leads to a reduction in these common measures of firm performance due to the spreading of capacity-constrained resources.

---

[16] For illustration purpose, we use the case when it is optimal for a firm to focus all capacity-constrained capabilities in one market *m*.

[17] In our model, profitability is simply $\frac{p_m s_m - \frac{r}{\gamma_m T} s_m}{p_m s_m} = 1 - \frac{\frac{r}{\gamma_m T}}{p_m} = 1 - \frac{1}{q}$. As a result, profitability uniformly increases with Tobin's *q*.

[18] Note that in this case we compare the Tobin's *q* values associated with these two strategies by examining their ratio, which not only provides a clearer expression but also conceptually equivalent to the comparison based on difference made in equation (14).

4.4. A cross-sectional analysis of the diversification discount

In the previous section, we compared the profit margin of the diversification strategy and the weighted average of two (hypothetical) focused strategies by the same firm -- the same firm focusing in either the initial or the new segment. The virtue of this method is that it uses the same firm with the same stock of capabilities and market conditions and therefore does not suffer from the issue that there might exist systematic differences between diversified firms and focused firms. In this section, we introduce competition in a cross-sectional setting and demonstrate that when relative demand conditions change, among a set of focused firms with heterogeneous capabilities, the more capable one(s) diversify and as a consequence under perform their competitors that remain focused.

In order to highlight the role of heterogeneous capabilities, we let the initial market condition be that there are n identical markets, each of which is a Bertrand duopoly and contains one focused firm. These n markets are identical in the sense that they have the same market size, $s_m = s$, the same willingness to pay, $w_m = w$, and capabilities developed in one market are perfectly fungible when applied to another market, i.e., $\delta = 0$. We then specify the initial degree of firm heterogeneity or capability distribution across firms consistent with such an n-market-n-focused-firm distribution in equilibrium. Without loss of generality, we assume that firms' capability levels take the following order[19]

$$T_1 > T_2 > \ldots > T_n > T_{n+1} \ldots \qquad (17)$$

Note that T_{n+1} acts as a potential entrant, the role of which will be explained shortly.

Given the assumption that each market is a Bertrand duopoly, it is straightforward to show that a sufficient and necessary condition for such an n-market-n-focused-firm distribution to be an equilibrium is that

$$T_1 < 2T_n \qquad (18)$$

[19] For simplicity, we assume that firms differ in their level of capacity constrained capabilities T but have the same scale free capabilities γ.

Given that the product markets are of equal size, the optimal allocation for a focused firm entering an additional market is to split its capacity-constrained capabilities evenly between these two equal-size markets. Therefore, for the most capable firm (firm 1) to be competitively viable in the least capable firm's (firm *n*) home market, it must be the case that $T_1 > 2T_n$. Intuitively, the initial conditions (Equation 18) require that firm heterogeneity, which can be broadly defined by the difference between firm 1 and firm *n*, should not be too large.

Note that while competition in the current Bertrand model does not affect the continuous allocation of capacity-constrained capabilities to a given market, it does determine whether a firm is viable to enter a new market. In particular, there are more than *n* firms in condition (17), but since there're only *n* markets, those firms that have capability level lower than T_n will stay outside. However, the firm with capabilities T_{n+1} (equivalent to cost level $c_{n+1} = \frac{r}{\gamma T_{n+1}}$) sets the price for each of the *n* markets, because whenever a firm sets price higher than $c_{n+1} = \frac{r}{\gamma T_{n+1}}$, firm (*n*+1) will enter.[20] Moreover, only the capability level of firm (*n*+1), but not the other firms with further lower capability levels, matters.

Starting with the above conditions, we introduce a discrete type of demand change that one new market, or market (*n*+1), exogenously emerges. Given such a demand change, there can be two possibilities.

Case I: When $T_1 > 2T_{n+1}$, or the capability level of firm 1 is higher than two times that of firm (*n*+1).

In equilibrium, firm 1, the most capable firm, diversifies into this (*n*+1)th market, while the (*n*+1)th firm still remains outside though its capability level effectively sets the price: $p = c_{n+1} = \frac{r}{\gamma T_{n+1}}$.

[20] This is to assume that all firms have a cost level lower than willingness to pay.

At the same time, all the other firms still remain focused. Consequently, the most capable firm becomes diversified while the other less capable competitors remain focused.

After diversification, the diversified firm, or Firm 1, has a profit margin

$$mgn_1 = \frac{2s(p - \frac{r}{\gamma \frac{1}{2} T_1})}{2sp} = \frac{p - \frac{r}{\gamma \frac{1}{2} T_1}}{p}$$

At the same time, firm *n* has a profit margin

$$mgn_n = \frac{p - \frac{r}{\gamma T_n}}{p}$$

According to initial condition (18), or $T_1 < 2T_n$, we have

$$mgn_1 = \frac{p - \frac{r}{\gamma \frac{1}{2} T_1}}{p} < mgn_n = \frac{p - \frac{r}{\gamma T_n}}{p} \quad (19)$$

Given condition (17), we have $mgn_1 < [mgn_n < mgn_{n-1} ... < mgn_2]$. Therefore, in such a case, we obtain the observed diversification discount, which can in turn be explained by our theory. The most capable firm diversifies and its less capable competitors stay focused; however, at the same time, the diversified firm incurs lower average returns than the group of focused firms due to the spreading of its capacity-constrained capabilities.

Case II: when $T_1 < 2T_{n+1}$, or the capability level of firm 1 is lower than two times that of firm (*n*+1).

In such a case, the (*n*+1)th firm will enter this new (*n*+1)th market and there will be (*n*+1) firms each of which focusing in one of the (*n*+1) markets. In some sense, the firm-market distribution does not change in the sense that all firms are still focused.

The above results can be summarized by the following proposition.

Proposition 6. *For a distribution of firms specified by condition (17) and (18), when there is an intermediate degree of firm heterogeneity, or when* $2T_{n+1} < T_1 < 2T_n$ *, a set of n focused firms respond to exogenous demand changes (characterized by the emergence of a new market opportunity) in such a way that, more capable firms diversify and under perform those focused firms in terms of average returns.*

Heterogeneity among firms must be sufficiently great that the more capable firm, even with the spreading of its capacity-constrained capabilities across markets, can still be competitively viable across the multiple markets in which it is competing. However, at the same time heterogeneity must be bounded such that even the least capable of the initial set of *n* firms is competitively viable in the market in which it is focused.

5. Discussion

Our model suggests an alternative self-selection mechanism that can account for the observation of a cross-sectional diversification discount (Lang and Stulz 1994; Berg and Ofek 1995). Firms with superior capabilities in a low value (existing market) context are more likely to choose to diversify, but at the same time incur lower average return, despite the fact that the diversification decision enhances the firm's total profits. Such results are demonstrated in two ways. Section 4.3 and Proposition 5 uses a one-firm analysis, which essentially sets up a natural experiment that controls for the stock of capabilities and demand conditions, to decompose the distinct mechanism through which the rational trade-off between total profits and profit margins arises from the fact that capacity-constrained capabilities must be allocated across applications. Section 4.4 and Proposition 6 explicitly treat a set of competing firms that differ in their level of capacity-constrained capabilities and demonstrate the existence of a cross-sectional diversification discount in equilibrium where, in their rational responses to exogenous demand changes, more capable firms choose to diversify but under perform their less capable competitors that stay focused. Therefore, it may not be, as suggested by Gomes and Livdan (2004), that those firms with less capabilities (low productivity) diversify first and this sorting of "bad types" into diversification events explains the observed cross-sectional diversification discount; rather, it could be that those firms with more capabilities diversify first and that this diversification activity decreases average returns.

In this sense, our model can help reconcile the conflict between the existing self-selection explanations that rely on the assumptions of comparative productivity differences and diminishing returns to production scale (low productivity firms diversify first) and the proposition well established in the strategy field that firms with more relevant capabilities tend to enter a new field earlier (e.g., Klepper and Simons 2000; Mitchell 1989). Critical to our argument is the opportunity cost of applying capacity-constrained capabilities in diminishing opportunities. Our argument suggests that diversifying firms are "good types" (i.e., high capabilities) operating in "bad" market contexts.

Along these lines, a cross-sectional diversification discount may arise when firms participate in distinct niches in the same broadly defined industry. Different firms may experience different degrees of market maturity and those operating in more mature sub-markets are more likely to diversify and do so earlier. Alternatively, a "generalist" firm (Hannan and Freeman, 1989) may respond to the demand maturity earlier by diversifying because it has greater exposure to the overall market conditions, while a "specialist" may not do so if its demand conditions are less affected by market maturity. In either case, such rationally diversifying firms suffer the triple blow of facing a less attractive demand environment with the decline in size of their original market, the diminished effectiveness of their capabilities as these capabilities are applied to related, but distinct, product markets, and the spreading of capacity-constrained capabilities across more segments.

The current theoretical model provides a conceptual basis for subsequent empirical analysis to sort out the different arguments regarding the self-selection mechanism in the diversification process. We make distinct empirical predictions from the existing corporate finance literature regarding which firms (more or less capable) are more or less diversified. Existing industry level studies, such as Klepper and Simons's (2000) work on the TV receiver industry, are broadly consistent with the arguments developed here. As commercial broadcasting began after World War II, the demand for TV receivers took off rapidly and attracted a flood of entrants (through 1989 a total of 177 US firms), many of which came from the radio industry. Klepper and Simons (2000) find that a greater degree of radio experience, measured by firm size, types of radios, and years of production, significantly increased the likelihood and speed of entry. Thus, radio producers appear to diversify into the TV receiver industry as a response to the growth of the TV market, and the relative maturity of the radio market; furthermore, if we interpret more experience as evidence of more capabilities, then the results suggest that firms with more capabilities tend to diversify earlier.

As a more general methodological note, it is worth observing that research on the relationship between prior experience/capabilities and entry based on a fined-grained industry classification is able to offer more refined measures of firms' skills and capabilities than cross-industry analysis that inevitably much rely on more coarse-grained data. Thus, the analysis that contrasts de novo entrants versus de alia entrants, such as Klepper and Simmons (2000), Carroll, Bigelow, Seidel, and Tsai (1996), and Helfat and Lieberman (2002), offers an important window to a capability based logic of diversification. Along these lines, more refined empirical analyses allow for measures of market demand that more closely correspond to the actual product market conditions. Even industry classification at the four-digit SIC level may incorporate many rather distinct submarkets with quite different demand patterns.[21] This more refined sort of empirical analysis appears necessary to further unpack the critical elements of firm heterogeneity which results in firms being "sorted" into diversification activity. Is the sorting into diversification activity based on exogenous market maturity and a high level of capacity-constrained capabilities that have lost their value in their current application as suggested here, or is the differential sorting into diversification driven by low levels of firm capabilities and associated relatively ex-ante weak performance as suggested by recent writings in the corporate finance literature (e.g., Gomes and Livdan 2004)?

---

[21] A good example is the cardiovascular medical device industry. Although this industry is underneath primary SIC 3845 (Electromedical and Electrotherapeutic Apparatus), the relevant demand conditions for a given manufacturer are far more nuanced than a four-digit measure would provide, because there exist multiple independent product sub-markets, such as stents, pacemakers, and heart valves, that have experienced very different industry life cycles (Wu 2007).

While the contemporary literature on diversification from a resource perspective builds upon Penrose's (1959) idea of excess firm capabilities, the emphasis has been focused on the fungibility of resources across domains. Making a clear analytical distinction between scale-free and capacity-constrained capabilities helps to shift the discourse back to some of Penrose's (1959) original interest in the stock of organizational capabilities. The existence of capacity-constrained capabilities implies that these capabilities must be allocated across applications and therefore rational diversification decisions should be based upon the opportunity cost of their use in one domain or another, which is in turn determined by the relative size of different market segments and the degree to which the effectiveness of scale-free capabilities diminishes across markets. We further identify the demand thresholds for firms to diversify or exit as functions of a firm's total amount of scale-free capabilities and capacity-constrained capabilities, which allows us to infer the effect of heterogeneous capabilities on the order of diversification. The recognition of capacity-constrained capabilities also provides a rational explanation for the divergence between total profits and profit margins and in turn an alternative explanation of the diversification discount.

Arguably, a failing of early treatments of the resource view of the firm within the strategy literature is that the work tended to focused nearly exclusively on imperfections in input markets and tended to ignore the role of product markets. As illustrated by Adner and Levinthal (2001), Adner and Zemskey (2006), and Peteraf and Barney (2003), it is important to consider both demand and supply considerations in examining the value of resources. Similarly, diversification is not merely driven by supply-side considerations of rare and distinctive resources, but is equally impacted by the market opportunities to which these resources may be applied. These ideas are not entirely new. In many respects, they were largely intimated by Penrose's seminal work. However, we hope to have provided some important new analytical insights on these questions --- insights that help link some of the current re-examination of the diversified firm to this earlier Penrosian tradition within the strategy literature.

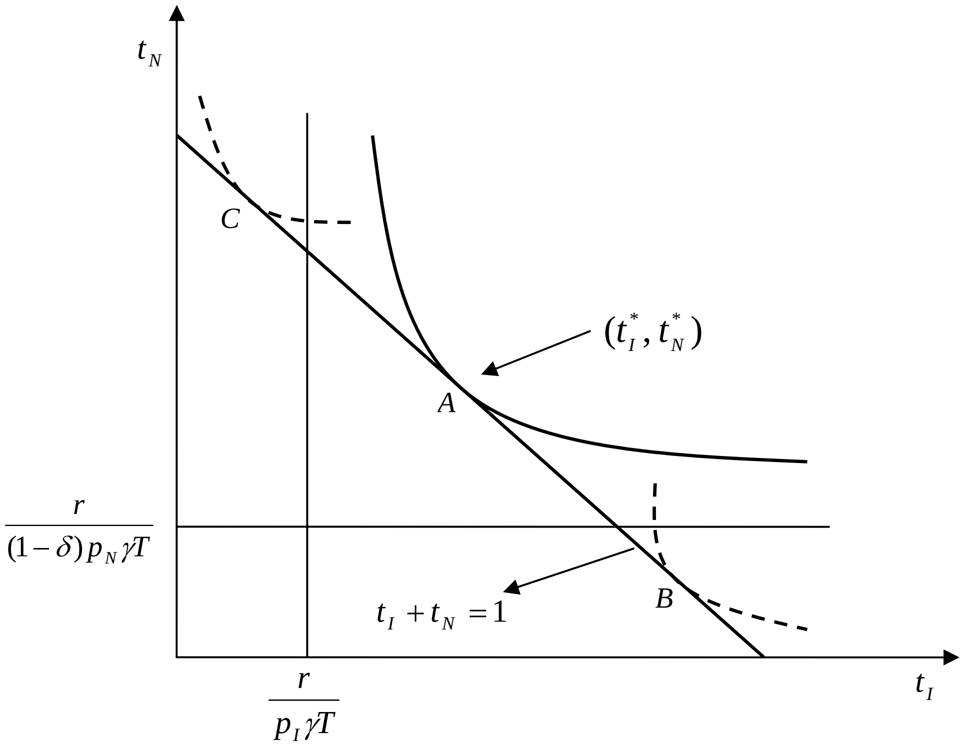

Figure 1: The allocation of capacity-constrained capabilities

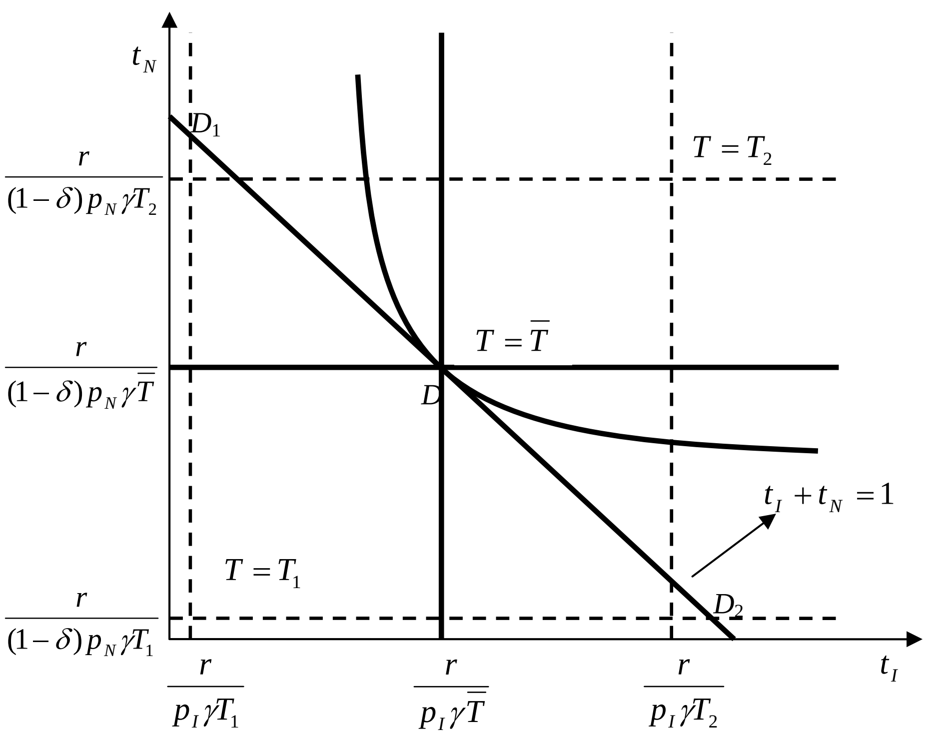

Figure 3: Total capabilities and the extent of diversification

Appendix 1:

Assume that there are *M* markets and market 1 is the initial market segment. A given firm has scale free capabilities denoted by γ. The factor of imperfect fungibility is assumed to be δ_m for each of the *M* segments, $m = 1, 2, \ldots, M$. Without loss of generality, let $0 = \delta_1 < \delta_2 < \ldots < \delta_M$, where $\delta_1 = 0$ means that the effectiveness of capabilities does not diminish in the original segment.

Maximization problem (5) can be extended to the case of *M* segments:

$$\min\{\sum_{m=1}^{M} \frac{s_m}{\gamma_m t_m T} \mid \sum_{m=1}^{M} t_m = 1, \frac{r}{\gamma_m t_m T} \le p_m\}$$

The fraction of capacity-constrained capabilities allocated to segment *m* is

$$t_m^* = \frac{\sqrt{\dfrac{s_m}{1-\delta_m}}}{\sum_{m=1}^{n} \sqrt{\dfrac{s_m}{1-\delta_m}}} = \frac{\sqrt{\dfrac{s_m}{1-\delta_m}}}{S_M}$$

where for notation simplicity, we denote $S_M = \sum_{m=1}^{M} \sqrt{\dfrac{s_m}{1-\delta_m}}$.

For the *m*th segment to be viable, the following condition must be satisfied

$$p_m - \frac{r}{\gamma_m t_m T} = p_m - \frac{r}{(1-\delta_m)\gamma T \dfrac{\sqrt{\dfrac{s_m}{1-\delta_m}}}{S_M}} > 0$$

Accordingly, we can derive the demand threshold for the *m*th segment to be viable

$$\frac{\sqrt{\dfrac{s_m}{1-\delta_m}}}{S_M} > \frac{r}{p_m (1-\delta_m)\gamma T}$$

where $m = 1, 2, \ldots, M$

Our purpose is to derive a capability threshold for γT , below which a firm cannot support M distinct segments simultaneously, under whatever demand conditions. Summing up the M necessary conditions leads to

$$1 = \frac{\sum_{m=1}^{M} \sqrt{\frac{s_m}{1-\delta_m}}}{S_M} > \frac{r}{\gamma T} \sum_{m=1}^{M} \frac{1}{p_m (1-\delta_m)}$$

Finally, we obtain the capability threshold for these M distinct segments to be viable simultaneously

$$\overline{T} = \frac{r}{\gamma} \sum_{m=1}^{M} \frac{1}{p_m (1-\delta_m)} \qquad (16)$$

For all these M segments to be viable, the firm must have more than $\overline{T}$ total capabilities. On the other hand, given total capabilities T , we can derive the maximal number of markets in which the firm can possibly participate. Equation (16) shows that the threshold is monotonically increasing with the number of segments M. Therefore, for a given stock of total capabilities, there exists an M^* such that M^* market segments are simultaneously viable but (M^*+1) market segments are not and M^* thereby indicates the maximum number of segments that can be supported. More capabilities can support more market segments, or result in a larger value of M^*. The case of M markets shows that the stock of total capabilities constrains the number of markets segments in which a firm can be engaged.

Appendix 2:

The Tobin's q associated with the diversification strategy is

$$q_{DIV} = \frac{p_I s_I + p_N s_N}{r(k_I^* + k_N^*)}$$

where capital investment in each market is respectively $k_I^* = \frac{s_I}{\gamma_I T t_I^*}$ and $k_N^* = \frac{s_I}{(1-\delta)\gamma_I T t_N^*}$.

Note that (i) if all capabilities are focused in the initial market, then capital investment is $k_I = \frac{s_I}{\gamma_I T}$; (ii) if all capabilities are focused in the new market, then capital investment is $k_N = \frac{s_I}{(1-\delta)\gamma_I T}$. Therefore, $k_I^* = \frac{1}{t_I^*} k_I$ and $k_N^* = \frac{1}{t_N^*} k_N$. Moreover, remember $t_I^* = \frac{\sqrt{(1-\delta)s_I}}{\sqrt{(1-\delta)s_I} + \sqrt{s_N}}$ and $t_N^* = \frac{\sqrt{s_N}}{\sqrt{(1-\delta)s_I} + \sqrt{s_N}}$, so we have $\frac{1}{t_I^*} = 1 + \frac{\sqrt{s_N}}{\sqrt{(1-\delta)s_I}}$ and $\frac{1}{t_N^*} = 1 + \frac{\sqrt{(1-\delta)s_I}}{\sqrt{s_N}}$.

Therefore, the Tobin's q associated with the diversification strategy can be transformed as

$$q_{DIV} = \frac{p_I s_I + p_N s_N}{r(k_I^* + k_N^*)} = \frac{p_I s_I + p_N s_N}{r(k_I \frac{1}{t_I^*} + k_N \frac{1}{t_N^*})}$$

$$= \frac{p_I s_I + p_N s_N}{r(k_I + k_N + k_I \frac{\sqrt{s_N}}{\sqrt{(1-\delta)s_I}} + k_N \frac{\sqrt{(1-\delta)s_I}}{\sqrt{s_N}})}$$

$$= \frac{p_I s_I + p_N s_N}{r(\frac{s_I}{\gamma_I T} + \frac{s_I}{(1-\delta)\gamma_I T} + \frac{s_I}{\gamma_I T} \frac{\sqrt{s_N}}{\sqrt{(1-\delta)s_I}} + \frac{s_N}{(1-\delta)\gamma_I T} \frac{\sqrt{(1-\delta)s_I}}{\sqrt{s_N}})}$$

$$= \frac{p_I s_I + p_N s_N}{\frac{r}{\gamma_I T}(s_I + \frac{s_N}{1-\delta} + 2\sqrt{\frac{s_I s_N}{1-\delta}})}$$

Next, we specify the weighted average of two focused strategies, with capital investment required when all capabilities are focused in one market, $k_I = \frac{s_I}{\gamma_I T}$ and $k_N = \frac{s_I}{(1-\delta)\gamma_I T}$, as weights, as

$$q_{FOC} = \frac{rk_I}{rk_I + rk_N} \frac{p_I s_I}{rk_I} + \frac{rk_N}{rk_I + rk_N} \frac{p_N s_N}{rk_N}$$

$$= \frac{p_I s_I + p_N s_N}{r(k_I + k_N)} = \frac{p_I s_I + p_N s_N}{\frac{r}{\gamma_I T}(s_I + \frac{s_N}{1-\delta})}$$

Finally, we compare the Tobin's *q* values associated with these two strategies by examining their ratio:

$$\frac{q_{FOC}}{q_{DIV}} = 1 + \frac{2\sqrt{\frac{s_I s_N}{1-\delta}}}{s_I + \frac{s_N}{1-\delta}}$$

Clearly, there exists a discount factor

$$\frac{2\sqrt{\frac{s_I s_N}{1-\delta}}}{s_I + \frac{s_N}{1-\delta}}$$

References

Adner, R. and D. Levinthal. 2001. Demand heterogeneity and technology evolution: Implications for product and process innovation. ***Management Science*** 47(5):611–628.

Adner, R. and P. Zemsky. 2006. A demand-based perspective on sustainable competitive advantage. ***Strategic Management Journal*** 27(3):215–239.

Anand, J. and H. Singh 1997. Asset redeployment, acquisitions and corporate strategy in declining industries. ***Strategic Management Journal*** 18(Summer): 99-118.

Barney, J.B. 1986. Strategic factor markets. ***Management Science*** 32(10): 1231-1241.

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

Campa, J.M. and S. Kedia, 2002. Explaining the diversification discount. ***Journal of Finance*** 57(4): 1731–1762.

Carroll, G.R., L.S. Bigelow, M.D.L. Seidel, and L.B. Tsai 1996. The fates of de novo and de alio producers in the American automobile industry 1885-1981. ***Strategic Management Journal*** 17(Summer Special Issue): 117-137.

Dierickx, I. and K. Cool. 1989. Asset stock accumulation and sustainability of competitive advantage. ***Management Science*** 35(12): 1504-1512.

Gomes, J. and D. Livdan 2004. Optimal diversification: Reconciling theory and evidence. ***The Journal of Finance*** 59 (2): 507-535.

Hannan, M. and J. Freeman. 1989. ***Organizational Ecology*.** Harvard University Press.

Helfat, C. 2003. Stylized facts regarding the evolution of organizational resources and capabilities. in C. Helfat, ed. 2003. ***The Blackwell/Strategic Management Society Handbook of Organizational Capabilities: Emergence, Development, and Change***, Blackwell Publishers.

Helfat, C. and M. Lieberman. 2002. The Birth of capabilities: Market entry and the importance of pre-history. ***Industrial and Corporate Change*** 11(4):725-760.

Helfat, C. and K.M. Eisenhardt. 2004 Inter-temporal economies of scope, organizational modularity, and the dynamics of diversification. ***Strategic Management Journal*** 25(13): 1217 – 1232

Jensen, M. 1986. Agency costs of free cash flow, corporate finance and takeovers. ***American Economic Review*** 76 (5): 323-329

Klepper, S. and K.L. Simons. 2000. Dominance by birthright: entry of prior radio producers and competitive ramifications in the U.S. television receiver industry. ***Strategic Management Journal*** 21(10-11): 997-1016.

Lang, L.H.P. and R.M. Stulz. 1994. Tobin's *q*, corporate diversification, and firm performance. ***Journal of Political Economy*** 102(6): 1248–1280.

Lindenberg, E.B. and S.A. Ross. 1981. Tobin's *q* ratio and industrial organization. ***Journal of Business*** 54(1): 1-32

Maksimovic, V. and G. Phillips. 2002. Do conglomerate firms allocate resources inefficiently across industries? Theory and evidence. ***Journal of Finance*** 57(2), 721-767.

Markides C.C. and P.J.Williamson. 1994. Related diversification, core competencies and corporate performance. ***Strategic Management Journal*** 15(special issue): 149-167

Mitchell, W. 1989. Whether and when? Probability and timing of incumbents/ entry into emerging industry subfields. ***Administrative Science Quarterly*** 34(2):208–230.

Montgomery, C. A. and B. Wernerfelt. 1988. Diversification, Ricardian Rents, and Tobin's *q*. ***RAND Journal of Economics*** 19(4): 623-632.

Palich, L. E., L. B. Cardinal, and C.C. Miller. 2000. Curvilinearity in the diversification-performance linkage: An examination of over three decades of research. ***Strategic Management Journal*** 21(2): 155-174.

Penrose, E. 1959. ***The Theory of the Growth of the Firm***, Oxford, Basil Blackwell.

Peteraf, M. A. and J. B. Barney. 2003. Unraveling the resource-based tangle. ***Managerial and Decision Economics*** 24(4): 309–323

Roberts, P.W. and S. McEvily. 2005. Product line expansion and resource cannibalization. ***Journal of Economic Behavior and Organization*** 57(1): 49-70.

Robins, J., and M. F. Wiersema. 1995. A resource-based approach to the multi-business firm: Empirical analysis of portfolio interrelationships and corporate financial performance. ***Strategic Management Journal*** 16(4): 277–299.

Rosen, S. 1982. Authority, control, and the distribution of earnings. ***The Bell Journal of Economics*** 13(2): 311-323.

Rubin, P. H. 1973. The expansion of firms. ***Journal of Political Economy*** 81(4): 936-49.

Rumelt R.P. 1974. ***Strategy, structure, and economic performance***. Boston, MA: Harvard Business School Press.

Schoar, A. 2002. Effects of corporate diversification on productivity. ***Journal of Finance*** 57(6): 2379-2403.

Slater, M. 1980. The managerial limitation to the growth of firms. ***Economic Journal***. 90(359): 520-528

Teece, D.J. 1980. Economies of scope and the scope of the enterprise. ***Journal of Economic Behavior and Organization*** 1(3): 223–247.

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

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

Winter, S.G. 1995. The four Rs of profitability: Rents, resources, routines, and replication. ***Resource-based and evolutionary theories of the firm: Towards a synthesis***. C. A. Montgomery. Boston, MA, Kluwer Academic Publishers: 147-178.

Winter, S.G. and G. Szulanski. 2001. Replication as strategy. ***Organization Science*** 12(6): 730–793.

Wu, B. 2007. Capacity-constrained capabilities, market maturity, and corporate diversification: Evidence from the cardiovascular medical device industry 1976-2004. Mack Center Working Paper, Wharton School.