7-2010

Opportunity Costs and Non-Scale Free Capabilities: Profit Maximization, Corporate Scope, and Profit margins

Daniel A. Levinthal
University of Pennsylvania

Brian Wu

Opportunity Costs and Non-Scale Free Capabilities: Profit Maximization, Corporate Scope, and Profit margins

Abstract
The resource-based view on firm diversification, subsequent to Penrose (1959), has focused primarily on the fungibility of resources across domains. We make a clear analytical distinction between scale free capabilities and those that are subject to opportunity costs and must be allocated to one use or another, thereby shifting the discourse back to Penrose's (1959) original argument regarding the stock of organizational capabilities. The existence of resources and capabilities that must be allocated across alternative uses implies that profit-maximizing diversification decisions should be based upon the opportunity cost of their use in one domain or another. This opportunity cost logic provides a rational explanation for the divergence between total profits and profit margins. Firms make profit-maximizing decisions to increase total profit via diversification when the industries in which they are currently competing become relatively mature. Due to the spreading of these capabilities across more segments, we may observe that firms' profit-maximizing diversification actions lead to total profit growth but lower average returns. The model provides an alternative explanation for empirical observations regarding the diversification discount. The self-selection effect noted in recent work in corporate finance may not be indicative of inferior capabilities of diversifying firms but of the limited opportunity contexts in which these firms are operating.

Keywords
corporate diversification, firm capabilities, opportunity cost, diversification discount, industry dynamics, resource-based view of the firm

Disciplines
Business Administration, Management, and Operations

Opportunity Costs and Non-Scale Free Capabilities: Profit Maximization, Corporate Scope, and Profit Margins

Daniel A. Levinthal*

Reginald H. Jones Professor of Corporate Strategy
3209 Steinberg-Dietrich Hall
Wharton School, University of Pennsylvania
Philadelphia, PA 19104-6370
215-898-6826 (phone)
215-573-8602 (fax)
dlev@wharton.upenn.edu

Brian Wu

Assistant Professor of Strategy
Stephen M. Ross School of Business
University of Michigan
701 Tappan Street, R4388
Ann Arbor, MI 48109-1234
734-647-9542 (phone)
734-764-2555 (fax)
wux@umich.edu

Forthcoming in ***Strategic Management Journal***

August 2009

* Corresponding author

We wish to thank the Mack Center for Emerging Technologies for generously supporting this research. We have benefited from comments on prior drafts by Ron Adner, David Collis, Glenn MacDonald, Costas Markides, Nicolaj Siggelkow, Harbir Singh, Sid Winter, the associate editor Joseph Mahoney, and two anonymous referees, as well as seminar audiences at the Tuck School, Dartmouth College, the Atlanta Competitive Advantage Conference, the Academy of Management, and the Harvard Strategy Conference.

Opportunity Costs and Non-Scale Free Capabilities: Profit Maximization, Corporate Scope, and Profit Margins

Abstract

The resource-based view on firm diversification, subsequent to Penrose (1959), has focused primarily on the fungibility of resources across domains. We make a clear analytical distinction between scale-free capabilities and those that are subject to opportunity costs and must be allocated to one use or another, thereby shifting the discourse back to Penrose's (1959) original argument regarding the stock of organizational capabilities. The existence of resources and capabilities that must be allocated across alternative uses implies that profit-maximizing diversification decisions should be based upon the opportunity cost of their use in one domain or another. This opportunity cost logic provides a rational explanation for the divergence between total profits and profit margins. Firms make profit-maximizing decisions to increase total profit via diversification when the industries in which they are currently competing become relatively mature. Due to the spreading of these capabilities across more segments, we may observe that firms' profit-maximizing diversification actions lead to total profit growth but lower average returns. The model provides an alternative explanation for empirical observations regarding the diversification discount. The self-selection effect noted in recent work in corporate finance may not be indicative of inferior capabilities of diversifying firms but of the limited opportunity contexts in which these firms are operating.

Running Head: Opportunity Costs and Non-Scale Free Capabilities

Keywords: corporate diversification, firm capabilities, opportunity cost, diversification discount, industry dynanmics, resource view of the firm

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). As Mahoney and Pandian (1992) note, this argument is based both upon the availability of resources, in particular the degree to which there may be slack resources in the firm's current market context, as well as implications of the nature of the firm's resources for the direction of possible diversification efforts. The diversification literature along the lines of the resource-based view has largely focused on this latter point highlighting the fungibility of resources, or 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 (cf., Montgomery and Wernerfelt, 1988). The fungibility of resources is the basis for the explanation as to why related diversification tends to outperform unrelated diversifications and, in turn, why firms tend to pursue more related diversifications (Bettis, 1981; Markides and Williamson, 1994; Montgomery and Wernerfelt, 1988; Robins and Wiersema, 1995; Rumelt, 1974).[2]

Implicitly, and at times explicitly, resources are often treated as having a *scale-free* property in the sense that the value of resources is assumed to be not reduced as a result of the sheer magnitude of firm operations over which they are applied. As Chang (1995: 387) notes, "The dominant view in diversification research is that intangible resources, such as technology and marketing skills, encourage firms to diversify into new businesses in order to exploit the 'public goods' nature of information-intensive assets." However, as argued in Penrose (1959),

[1] "Resources" and "capabilities" are used interchangeably in this paper.

[2] It is also important to note that owning a more fungible resource does not necessarily lead to competitive advantage, since it might be in more abundant supply (Montgomery and Wernerfelt, 1988) or attract more competition (Adner and Zemsky, 2008).

the stock of a firm's resources and the degree to which they are fungible across product markets are both critical in determining diversification decisions. Many of the resources that may underpin 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 subject to opportunity costs. 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. While some resources, such as a brand name or patent, may have a public good-like quality, most firm resources or capabilities do not.[3] The most familiar example in the business setting is a firm-specific management team (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).

This issue of the need to allocate capabilities across markets and, as a consequence, the linking of diversification efforts to demand conditions in alternative markets is highlighted by Chandler (1969), which maintains that it was the decline in product market activity in the late 1920s and the 1930s that precipitated the enormous growth in diversification of industrial firms in the U.S. Chandler (1969: 275) submits that firms such as DuPont and General Electric that "had accumulated vast resources in skilled manpower, facilities, and equipment" were under great pressure to find new markets as their existing ones ceased to grow.

More generally, as Mahoney and Pandian (1992) point out, there is an important line of inquiry running from Uzawa (1969), Chandler (1969; 1977), Rubin (1973), Slater (1980), and Teece (1982) that takes onboard Penrose's (1959) concern for the dynamics of resource accumulation by the firm and the implications of these dynamics for diversification efforts. In a

[3] Even a resource such as a brand name may not be a pure public good as the application of a brand name for one product might impact its value in another. Thus, when Gucci wildly applied its brand name to a number of lower-end products in the 1980s, the value of the brand was argued to have been reduced (Aaker, 2004).

similar vein, Helfat and Eisenhardt (2004) suggest that firms may reallocate resources across domains over time in order to achieve, what they term, inter-temporal scope economies.

However, the research literature has not been clear about the distinction between scale-free capabilities and those capabilities that must be allocated among alternative uses. This distinction is critical because we believe that it is this latter class of capabilities that capture the essence of Penrose's (1959) 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, since scale-free capabilities can always be leveraged in other areas and hence will always have "excess" capacity. Thus, it is only resources subject to an opportunity cost that affect how resources should optimally be allocated.

In this sense, we are making an analytical return to the original sensibility of Penrose's (1959) capability-based perspective on diversification. However, it is important to note that Penrose (1959) focused on the limit case in which firms had "excess" capabilities. This excess derives from two possible sources. One stems from the fact that some resources constitute discrete investments, such as a physical plant. If a manufacturing facility is not being fully utilized in the production of the firm's current products, then such a facility can be a free resource that can be applied, in part, to other means. The other source results from managerial learning, which enables a given managerial team to handle a greater range of responsibilities over time. Yet, resources, particularly human capital-based resources, tend to be fully allocated to particular tasks and initiatives at any point in time. In that sense, such resources are not in "excess." However, there still remains the question of what is the best allocation of the time of a sales force, product development team, or top management group based on their opportunity costs.

In addition to extending Penrose's (1959) notion of excess resources to the more general notion of opportunity cost, we incorporate a greater consideration of the role of the demand environment in influencing the opportunity costs associated with a firm's resources. The emphasis in Penrose's (1959) work was 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. 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 are based upon the opportunity cost of applying capabilities in one domain or another (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 could either reduce the value of applying non-scale free capabilities in the current market or raise the opportunity cost of not applying some of these capabilities in related product markets.[4] It is in this sense that resources become "underutilized" or "excess." Alternatively, if the current market continues to offer sufficiently favorable opportunities, it will not be economically rational to divert non-scale free resources 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

[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.

explanation of firms' diversification behavior in trading off profit margins for corporate growth. Largely ignored by the research literature is the fact that profit-maximizing diversification decisions imply that firms seek to increase total profit but not necessarily their profit margin or market-to-book value, with the latter 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 non-scale free 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' profit-maximizing diversification actions lead to total profit growth but lower average returns. In a similar vein, Montgomery and 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 capabilities to new product markets, even in the absence of any imperfect fungibility of firm-specific capabilities.

We first examine this model in a simple Bertrand set-up in which we demonstrate the basic result regarding the implication of profit maximizing diversification on profit margins. We then expand this to a Cournot model that allows for a more explicit treatment of competition. In the Cournot setting, we can generate cross-sectional results in which a firm with inferior capabilities remains focused on a single product market and earns higher profit margins, but less profit, than its more capable competitor. We also find that as firms become more asymmetric in their capabilities, the necessary size of the new product market to elicit diversification on the part of the more capable firm rises. With greater asymmetry in capabilities, the original market in which both firms compete becomes more attractive, thereby diminishing the incentive to

diversify. Thus, the Cournot results allow us to expand the considerations of competitive effects and also demonstrate the robustness of our original results to those effects.

The remainder of the paper is organized as follows. The next section further develops the notion of non-scale free capabilities that must be allocated among alternative uses and its contrast with scale-free capabilities. We then set up and analyze the formal model by linking capability-based arguments regarding diversification and the demand conditions in the markets in which the firm does and may participate. After developing some general results under Bertrand and Cournot competition for the relationship among diversification decisions, firms' capabilities, and profit margins, we engage in a numeric analysis of the Cournot model to further develop the implications of competitive forces on both diversification decisions and profit margins. Finally, we discuss the broader implications of these results.

OPPORTUNITY COST AND DIVERSIFICATION

Insert Table 1 about here

Table 1 illustrates the contrast between what we term scale-free capabilities and those resources that are congestible and require allocation to distinct purposes. In addition to this dimension by which capabilities may differ, there is the more traditional issue of fungibility, or the range of activities over which a resource or capability may be applied. The most restricted sort of resources resides in the lower left cell. Highly specific human or physical capital not only has the property that its specificity narrows the domain of activities over which the resource can be applied, but also that its use in one activity constrains its possible use in other activities. The cell in the upper left again offers human and physical capital examples, but in this case the capital is applicable over a wider domain of activities. While an auditor may not be usefully

applied in a marketing function, he or she can apply their auditing expertise to a wide variety of enterprises. Similarly, power generation equipment is a general purpose capital infrastructure that could be applied to an enormous range of uses. However, per the issue of opportunity cost, it is important to note that the use of an auditor in one engagement restricts their possible use in another; and the application of a certain magnitude of kilowatts to one purpose reduces the kilowatts available for another. In the right-hand side, the resources are not subject to opportunity cost. There are no inherent constraints on how many goods or services can bear a common brand name. Nor does the use of a computer program on one machine preclude its use on another.[5] While the fungibility of a brand name or computer operating system is not limitless, the range of application is generally greater than either a specific piece of intellectual property or the range of uses of a particular customer relationship.

The extant research literature on diversification has tended to focus 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: 226). The recognition of scale-free capabilities has had a profound influence on both academic research and industry practice, since it highlights the role of knowledge and competence as strategic assets (Winter, 1987). Winter and Szulanski's (2001) study of replication processes provides a paradigmatic example of a scale-free capability, defining the Arrow core as the informational endowment a firm extracts from an original setting which can be replicated to other settings. The distinctive property of such information-like resource is that "unlike any resource that is rivalrous in use, an

[5] The producer may restrict the replication of the program through site licenses or copy restrictions but clearly such constraints are to address issues of value appropriation and are not related to technological constraints to the application of a given program to multiple applications.

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: 741).

The critical constraint in the application of scale-free resources is not by definition the scale of the operations over which they are applied, but rather the scope or range of their applicability. Indeed, the issue of resource relatedness and fungibility is arguably the most studied question in corporate strategy. Rumelt (1974), in a pioneering examination of this issue, showed that firms pursuing related diversifications outperform those pursuing a strategy of unrelated diversification. This basic finding has been reconsidered with a variety of different measures of relatedness, but the general result has stood up (Bettis, 1981; Markides and Williamson, 1994; Montgomery and Wernerfelt, 1988; Robins and Wiersema, 1995).

While Penrose's (1959) emphasis on a firm's stock of capabilities as a basis for diversification has played a secondary role to the consideration of the fungibility of resources, there has been some awareness in the literature that there may be opportunity costs associated with the use of resources. In his brief discussion on the limits to diversification economies, Teece (1982: 53) suggests that, "Know-how is generally not embodied in blueprints alone; the human factor is critically important in technology transfer. Accordingly, as the demands for sharing know-how increase, bottlenecks in the form of over-extended scientists, engineers, and managers can be anticipated." Recent empirical work in both finance and management has provided suggestive evidence of the opportunity cost of allocating resources from one product domain to another. Schoar (2002) finds that after a firm diversifies into a new industry by acquiring a plant, the incumbent plants will incur a decrease in productivity, while the acquired plants increase productivity. Similarly, Roberts and McEvily (2005) find that entering a new pharmaceutical product market reduces a firm's performance in its current markets. Finally, Hitt *et al.* (1991:

695) observe that acquisitions tend to reduce both the extent and productivity of firms' R&D investments, suggesting that the "resources remaining for managerial allocation may become constrained, causing managers to forgo other investment opportunities."

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 mobile phones and consumer electronics. 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, on the whole, Intel's diversification efforts will increase sales and total profit. However, 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.[6]

Based upon the above reasoning, we develop the following arguments regarding diversification efforts. First, it is important to distinguish between diversification efforts based on scale-free and non-scale free capabilities. A scale-free resource, such as brand name, faces limits on the breadth of its fungibility (i.e., how broadly fungible is a given brand name) but not on its extent of application (i.e., the number of markets in which a given brand can be applied for a given level of fungibility). In contrast, the application of those non-scale free capabilities is

[6] Of course, another basis for the decline in average return is the shift from a market in which Intel has a dominant 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 non-scale free capabilities.

driven by the logic of opportunity costs. On the margin, is the greatest value of these firm-specific capabilities 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. Thus, when there are multiple segments, the range of diversification activity is constrained by the total stock of capabilities. This analysis 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 into the diversification discount, the observation that diversified firms tend to have a lower valuation, typically measured as the relationship between market value to book value or Tobin's *q*, than an amalgamation of an equivalent set of focused firms (Berger and Ofek, 1995; Lang and Stulz, 1994). Agency theorists suggest that diversification destroys value for reasons such as managers' empire building behavior that aims to increase 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: 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 or marketing capabilities) tend to enter a new field earlier and perform better (e.g., Helfat, 2003; Klepper and Simons, 2000; Mitchell, 1989). 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 than a statement about the firm's relative lack of capabilities.[7] In the pursuit of the best use of the firm's non-scale free resources, there is some allocation of

[7] The recent treatments of the diversification discount such as Campa and Kedia (2002) and Villalonga (2004) do control for industry demand conditions when examining the self-selection effect of the act of diversification on the so-called diversification discount. However, due to data limitations, such work must rely on measures of industry at the four-digit level of the SIC code. This level of aggregation masks considerable diversity of demand environments at the product level. We elaborate on this issue in the discussion section.

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.

MODEL STRUCTURE

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 $m = N$). Production in each segment is described as $Q_m = \gamma_m t_m T k_m$, where γ_m is the firm's scale-free capabilities, t_m is the share of the firm's total non-scale free capabilities, T, that must be divided among activities, and capital k_m whose replacement cost per unit is r which reflects the current market value of capital. For the sake of simplicity, we assume that firms are endowed with a particular capability stock T and do not consider the cost of developing this, but only the opportunity cost of how it is applied across different market segments within the firm. 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 $\frac{rQ_m}{\gamma_m t_m T} = c_m \times Q_m$, where

$$c_m \equiv \frac{r}{\gamma_m t_m T} \tag{1}$$

Note that (i) scale-free capabilities, γ_m, and non-scale free capabilities, T, are firm specific and subject to imperfect input markets (Teece, 1982); (ii) providing more of capabilities reduces both total and marginal cost since it substitutes for the purchased capital; (iii) c_m is decreasing in t_m and increases to infinity as t_m approaches zero; (iv) we specify the following

relationship between scale-free capabilities in the two markets: $\gamma_N = (1-\delta)\gamma_I$. This relationship implies that scale-free capabilities, γ_I, are not perfectly fungible, and 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*.[8]

It is important to relate this production function and associated cost function to our underlying argument regarding scale-free and non-scale free capabilities. Our focus is on the allocation of non-scale-free capabilities across market contexts. This capability could be a product development team that might be spread across multiple initiatives, as highlighted in Christensen and Bower's (1996) work on the disk drive industry, the allocation of scarce production capacity as in Burgleman's (1994) work on Intel, or the attention of the top management team (Rosen, 1982). For the sake of simplicity, we are not addressing this allocation with respect to the scale of activity within a given business unit. Thus, a business unit is assumed to be able to scale up its activity at a constant marginal cost. This simplification follows on the works of Klepper (1996) and Lippman and Rumelt (1982) on industry evolution, which, in order to highlight the effect of across-firm heterogeneity, postulate a production technology with a constant marginal cost of production. This assumption of constant return to production scale also allows us to demonstrate that diversification can arise from the change in demand conditions in the absence of diminishing return to production scale, in contrast to recent work in corporate finance (cf., Gomes and Livdan, 2004; Maksimovic and Phillips, 2002) in which diminishing returns to production scale acts as an underlying driver for diversification.

[8] As we point out in Table 1, non-scale free capabilities are also subject to the issue of imperfect fungibility, but the distinctive feature of non-scale free capabilities is that they must be allocated across alternative uses based on opportunity costs. In order to highlight the issue of allocation, we assume that non-scale free 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.

We denote the demand and price for a given firm's product in segment m as q_m and p_m respectively. Should the firm engage in both activities, its profit is $q_I(p_I - c_I) + q_N(p_N - c_N)$ or $q_I(p_I - \frac{r}{\gamma_I t_I T}) + q_N(p_N - \frac{r}{\gamma_N t_N T})$. Assuming the optimal t_I and t_N are both strictly positive, the firm's problem is:

$$\max\{(p_I q_I + p_N q_N) - r(\frac{q_I}{\gamma_I t_I T} + \frac{q_N}{\gamma_N t_N T}) \mid t_I + t_N = 1, \frac{r}{\gamma_N t_N T} \le p_N, \text{and } \frac{r}{\gamma_I t_I T} \le p_I\} \quad (2)$$

Note that it may not be optimal for the firm to engage in both activities simultaneously. Furthermore, if it is optimal for the firm to be diversified into both activities, then the optimal allocations of non-scale-free capabilities t_I^* and t_N^* are bounded away from zero. That is, to get into the new activity requires a discrete reduction in the capabilities employed in the initial activity. Indeed, depending on the magnitude of the firm's capabilities and the market price, a firm may not be competitively viable in a given market.

This above setting offers a general framework for analyzing firms' diversification decisions and the associated performance effects based on (non-scale-free) capability allocation. In the following sections, we examine the diversification problem characterized in this general framework using the two most widely used market models: the Bertrand model and the Cournot model, which usefully complement each other. The Bertrand analysis allows us to develop the basic insights regarding the allocation of scarce capabilities across lines of business and its implications for firm profitability. Introducing Cournot competition allows us to examine more directly the impact of competitive interaction among firms. However, the Cournot analysis in this context does not lend itself to a full closed form analytical solution as the reaction functions are neither continuous nor monotonic. Thus, we analytically derive the general results regarding firm diversification decisions, and then perform a numerical analysis to characterize more

specific properties regarding profit margins and the pattern of diversification. Further, we analyze a wide array of parameter values to both explore further insights of the model and the robustness of our results.

ANALYSIS: Bertrand Model

Each market is specified to be a Bertrand duopoly, where whenever the firm allocates non-scale-free capabilities such that its marginal cost is lower than that of the competitor, the firm serves the whole market with demand. The demand side of segment m consists of s_m consumers. Thus, the firm that captures the market produces quantity $q_m = s_m$, charges a price of p_m equal to the competitor's marginal cost, and has profits $s_m(p_m - c_m)$, where $c_m = \frac{r}{\gamma_m t_m T}$ is defined as in equation (1).

In this Bertrand setting, a given firm diversifies into the new segment when demand conditions are such that the maximization problem (2) has an interior solution (t_I^*, t_N^*). The total profit associated with the diversification strategy is $s_I(p_I - \frac{r}{\gamma_I t_I^* T}) + s_N(p_N - \frac{r}{\gamma_N t_N^* T})$, andthe total sales associated with the diversification strategy is $p_I s_I + p_N s_N$. Therefore, the profit margin associated with the diversification strategy, which we denote as π^*, is

$$\pi^* = \frac{s_I}{p_I s_I + p_N s_N}(p_I - \frac{r}{\gamma_I T t_I^*}) + \frac{s_N}{p_I s_I + p_N s_N}(p_N - \frac{r}{(1-\delta)\gamma_I T t_N^*}) \qquad (3)$$

In contrast, when firms focus on the initial segment, the profit margin is $\frac{s_I(p_I - \frac{r}{\gamma_I T})}{p_I s_I}$, while 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, which we denote as π^w, 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

$$\pi^w = \frac{s_I}{p_I s_I + p_N s_N}(p_I - \frac{r}{\gamma_I T}) + \frac{s_N}{p_I s_I + p_N s_N}(p_N - \frac{r}{(1-\delta)\gamma_I T}) \quad (4)$$

π^w is a standard benchmark used to compare the performance of diversified and focused firms (Berger and Ofek, 1995; Lang and Stulz, 1994). The relation between the profit margin of the diversification strategy (equation (3)) and that of focused strategies (equation (4)) is characterized by equation (5).

$$\pi^* = \pi^w - 2\frac{\sqrt{s_I s_N}}{p_I s_I + p_N s_N}\frac{r}{\sqrt{(1-\delta)}\gamma_I T} \quad (5)$$

As a result, we have the following proposition.

Proposition 1: The profit margin of the profit-maximizing diversification strategy for a firm with capability stock T is lower than the weighted average of the profit margin of focusing this capability stock T in each of the two markets. The difference is characterized by equation (5).

(See Appendix 1 for a proof.)

Proposition 1 allows us to identify the sources of the declining profit margin associated with diversification by decomposing 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 non-scale-free capabilities. The second term of π^w in equation (4) indicates how 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 Tobin's q as firms apply, what we term, scale-free capabilities in increasingly distant markets.

Adding to this consideration of imperfect fungibility, the second term of equation (5) captures the discount due to the spreading of non-scale-free capabilities. Therefore, equation (5) provides a more complete picture of the diversification discount from the resource-based view by incorporating the effect of non-scale-free resources that need to be allocated across applications. Proposition 1 suggests that the existence of a diversification discount does not necessarily result from agency behavior that deviates from profit maximization. The spreading of non-scale-free capabilities across more applications based on opportunity costs implies that the profit margin will be "sacrificed" to some extent in the pursuit of total profit maximization.

In parallel to this discussion of the impact of diversification on the firm's profit margins, we can also show that rational diversification efforts lead to lower Tobin's q, a widely used measure in the empirical analysis of diversification performance. Given that there is no short-run variable production cost (e.g., depreciation and labor) in the model, the market value of the firm in this stylized one-period model can be represented by operating profit $p_m s_m$, which is the earnings stream that is generated from the firm's capital stock of $r\frac{s_m}{\gamma_m T}$. Therefore, Tobin's q, defined as the market value of the firm divided by the replacement cost of capital (Lindenberg and Ross, 1981; Winter, 1995), can be represented as $q = \frac{p_m s_m}{r\frac{s_m}{\gamma_m T}}$.

We denote the Tobin's q associated with the diversification strategy as q_{DIV} and the Tobin's q associated with the weighted average of two focused strategies as q_{FOC}. Analogous to the relationship in equation (5), the ratio of q_{FOC} and q_{DIV} is

$$\frac{q_{FOC}}{q_{DIV}} = 1 + \frac{2\sqrt{\frac{s_I s_N}{1-\delta}}}{s_I + \frac{s_N}{1-\delta}} \tag{6}$$

As a result, we have the following corollary.

Corollary 1: The Tobin's q of the profit-maximizing diversification strategy for a firm with capability stock T is lower than the weighted average of the Tobin's q of focusing this capability stock T in each of the two markets. The difference, in ratio terms, is characterized by equation (6).

(See Appendix 2 for a proof.)

Therefore, whether the average return to capital is measured as economic profit margin (equation (5)) or as Tobin's q (equation (6)), we see that profit-maximizing diversification leads to a reduction in these common measures of firm performance due to the spreading of firm resources over multiple product markets.

ANALYSIS: Cournot Model

The prior analysis examined the profit maximizing tradeoff between total profits and profit margins based on a Bertrand model. This analysis models the diversification of a focal firm that faces competitive constraints, but does not explicitly model an endogenous competitive process. In this section, we address this gap by introducing a model of Cournot competition. We develop this analysis using the same analytical structure for firms' capabilities and their associated cost functions, but the Cournot structure requires us to more fully specify a demand function.

Demand in the initial market is specified as: $p_I = a_I - b_I(q_{1I} + q_{2I})$, where q_{1I} and q_{2I} are the output of firms 1 and 2 in the initial market. Similarly, in the new market, demand is: $p_N = a_N - b_N(q_{1N} + q_{2N})$, where q_{1N} and q_{2N} are the output of firms 1 and 2 in the new market.

As before (equation (1)), firms are characterized by their non-scale-free capability T_1 and T_2, respectively. However, we set the value of scale-free capabilities γ for the two competitors to be the same as we are focusing on the role of non-scale-free capabilities. The marginal production costs for firms 1 and 2 in market I and N are given by $c_{1I} = \frac{1}{\gamma t_{1I} T_1}$, $c_{2I} = \frac{1}{\gamma t_{2I} T_2}$, $c_{1N} = \frac{1}{(1-\delta)\gamma(1-t_{1I})T_1}$, and $c_{2N} = \frac{1}{(1-\delta)\gamma(1-t_{2I})T_2}$, where t_{1I} and t_{2I} are the fraction of non-scale-free capability of firms 1 and 2 invested on the initial market, respectively, and $(1-\delta)$ is the effectiveness of scale-free capabilities, γ, when γ is applied to the new segment N ($0 \le \delta < 1$). Thus, firms' profits in each market are given by $\pi_{im} = (p_m - c_{im})q_{im}$, for $i = 1,2; m = I, N$.

The sequencing of the firms' choice problem is that firms first choose their non-scale-free capability allocations, characterized by t_{1I} and t_{2I}. Second, given the chosen t_{1I} and t_{2I}, firms choose their production quantities q_{1I} and q_{2I} in the initial market and q_{1N} and q_{2N} in the new market. The associated optimization problem can be well specified via backward induction. However, in this context, the problem does not lend itself to a general analytical solution. Indeed, we will show in the subsequent analysis that the best-response functions that underpin the equilibrium analysis may, in some cases, be non-continuous and non-monotonic.[9]

[9] In this respect, it is worth noting that our model is related to but distinct from a classic IO model developed by Kreps and Scheinkman (1983), which showed that a two-stage model of production capacity investment followed by Bertrand price competition can be treated in a relatively straightforward manner as being equivalent to a one-stage

The best response curves are not continuous for two reasons. First, the optimal allocations of non-scale-free capabilities are bounded away from zero, i.e., to get into the new activity requires a discrete reduction in the capabilities employed in the initial activity. It would never be optimal for a firm to allocate less than some epsilon to the new market. The firm must allocate sufficient capabilities into the new market such that it can establish competitive viability in this market. Furthermore, the firm must make sure the gains in the new market to which these capabilities are being applied is larger than the losses incurred by allocating these capabilities away from the old market. As a result, the best response curve always has a jump from zero at the point of diversification to some discrete magnitude.

The second reason that the best response curves may not be continuous is that the best response curves have small jumps at the point where the competitor either makes a discrete move into or out of a product market in response to its competitor's change in allocation of capabilities across markets. Note that this second reason for the best response curves not being continuous is also the reason why the best response curves are not monotonic (see Appendix 3 for an illustration of the above argument and some more detailed explanation).

Even in the presence of the analytical challenges identified above, we are still able to derive the following propositions. To capture the change in relative demand, without loss of

---

model of Cournot competition. The first-stage decision variable is production capacity in Kreps and Scheinkman's model, but in the context of our model it is the allocation of capabilities. The allocated capabilities determine the level of average cost in a given market (see equation (2) on p. 13). Also, in contrast to Kreps and Scheinkman (1983), we assume there is no constraint to production capacity and thus production capacity can scale up simultaneously with the output decision. Under this specification, the sequencing of the firms' choice problem is that firms first choose their non-scale free capability allocations, which in turn determine firms' cost level in each market. Second, given the chosen capability allocations and the associated cost levels, firms either choose their prices, if under the Bertrand competition, or choose production quantities, if under Cournot competition. The firm's optimization problem can then be analyzed via backward induction. In sum, our model differs from Kreps and Scheinkman's model in two important ways. First, Kreps and Scheinkman's (1983) model examines only one market, while we model the allocation of capabilities across multiple markets along with diversification decisions. Second, Kreps and Scheinkman's (1983) model examines Cournot competition assuming homogeneous firm capabilities, while our model examines how capabilities in each market are endogenously determined prior to Cournot competition.

generality, we hold constant the size of the initial market a_I and only vary the size of the new market a_N. Let $P_1 \triangleq P(a_N, T_1, T_2, t_{1I}, t_{2I})$ be firm 1's profit function given the new market size a_N, firm 1's capability T_1, the competitor firm 2's capability T_2, firm 1's allocation t_{1I}, and the competitor firm 2's allocation t_{2I}. Similarly, $P_2 \triangleq P(a_N, T_2, T_1, t_{2I}, t_{1I})$ represents firm 2's profit function given the new market size a_N, firm 2's capability T_2, the competitor firm 1's capability T_1, firm 2's allocation t_{2I}, and the competitor firm 1's allocation t_{1I}. We use t_{1I}^* and t_{2I}^* to denote firms' allocation in equilibrium. Given an initial setting ($a_N = 0$) in which both firms compete in the initial market, we are able to state the following propositions:

Proposition 2a: For $i = 1, 2$, there exists an $\hat{a}_N > 0, \eta > 0,$ and $\varepsilon > 0$, such that, when the size of the new market a_N gets sufficiently large but not too large ($\hat{a}_N < a_N < \hat{a}_N + \eta$) and the two firms' capability asymmetry is small enough ($\left|T_i - T_{(3-i)}\right| < \varepsilon$), there exists a pure strategy Nash equilibrium in which either firm may first diversify ($t_{(3-i)I}^* < 1$) while the other firm stays focused at the initial market ($t_{iI}^* = 1$).

Proposition 2b: For $i = 1, 2$, fixing T_i as the less capable firm's capability level, there exists a $\tilde{T} > 0$, $\hat{a}_N > 0$, and $\eta > 0$ such that, when the size of the new market a_N gets sufficiently large ($\hat{a}_N < a_N < \hat{a}_N + \eta$) and when the two firms' capability asymmetry is sufficiently large ($T_{(3-i)} > \tilde{T}$), there exists a unique pure strategy Nash equilibrium in which only the more capable firm will first diversify ($0 < t_{(3-i)I}^* < 1$) while the less capable firm stays focused at the initial market ($t_{iI}^* = 1$).

Proposition 2c: For $i = 1, 2$, given a_I as the size of the initial market and T_{3-i} as the more capable firm's capability level, there exists $\underline{T} > 0$ such that the firm without sufficient capabilities ($0 < T_i < \underline{T}$) will either focus on the initial market ($t_{iI}^* = 1$) or switch all resources from one market to the other ($t_{iI}^* = 0$) in any pure strategy Nash equilibrium.

(See Appendix 4 for proofs.)

These three propositions highlight how firms' diversification decisions differ when the size of the new market increases, depending on their capability asymmetry. Proposition 2a

highlights the role of competitive conditions being endogenous on the reallocation of resources across product markets. In this case, when one firm diversifies to take advantage of the new market opportunity, the current market becomes more attractive as the diversifying firm withdraws capabilities from this market. Therefore, there exists an equilibrium in which either the more capable firm or the less capable diversifies first while the other firm stays focused on the initial market.[10] In addition, it is not an equilibrium for both firms to start diversifying simultaneously as the relative demand changes, even when these two firms have equal capabilities.

However, while Proposition 2a provides a baseline result when firms are relatively symmetric, the core issue that is of primary interest to the strategy field is when firms are, to a significant degree, asymmetric in their capabilities. Proposition 2b demonstrates that with sufficient asymmetry in capabilities, there exists a unique pure strategy equilibrium in which the more capable firm diversifies first. In this case, for the more capable firm, the need to reallocate capabilities to the new market due to opportunity costs dominates the strategic consideration of waiting for the less capable firm to diversify first and thus alleviate competition in the initial segment.

Proposition 2(c) indicates that a minimum amount of capabilities are necessary for the diversification strategy ever to be the optimal strategy, irrespective of relative demand conditions across alternative markets. When relative demand changes, the more capable firm goes through a diversification process; however, without sufficient capabilities, the less capable firm may never become diversified. It will focus on the initial market when the relative market size of the new

[10] There could be two pure strategy Nash equilibria in Proposition 2a ($(t_{iI}^* = 1, 0 < t_{(3-i)I}^* < 1)$ or $(t_{iI}^* = 1, t_{(3-i)I}^* = 0)$). However, in either case, Proposition 2a holds. Either firm can diversify first (diversify partly into the new market $(0 < t_{(3-i)I}^* < 1)$ or completely switch to the new market $(t_{(3-i)I}^* = 0)$), while the other firm stays focused in the initial market $(t_{iI}^* = 1)$.

market is small. When the relative market size of the new market gets larger, the less capable firm may completely switch to the new market but will never simultaneously engage both markets. This proposition further highlights the critical role of non-scale-free resources in determining the scope of the firm.

In the following numerical analysis, we examine more fully the impact of asymmetry of capabilities on the diversification decisions and, in particular, the threshold of relative size of the new and the initial market that triggers diversification. In addition, we examine the impact on profit margins of diversification. This analysis not only generalizes Proposition 1 to the Cournot case, but it allows us to examine the cross-sectional implications of diversification behavior on the profit margins for firms with varying capability levels.

NUMERICAL ANALYSIS OF COMPETITIVE INTERACTION

For our numerical analysis, we set the baseline parameter values as follows: $a_I = 10$, $b_I = b_N = 1$, scale-free capabilities $\gamma = 1$, and the fungibility of scale-free capabilities $(1-\delta) = 0.8$. In the subsequent analysis, we highlight the impact of varying levels of firm capabilities and market size (a_N, the size of the new market). This later variable allows us to capture relative demand maturity based on the increase in the size of the new market a_N.[11] Holding constant the new market and varying the initial market would lead to the same results.

Implications of numerical investigation for diversification decisions

In the Cournot analysis, we are able to analytically show that when two firms' capability asymmetry is sufficiently large and the new market size is of sufficient size, the more capable

[11] Note that while we keep scale-free capabilities γ and their fungibility $(1-\delta)$ in the model to stay connected with the existing literature, we do not explore these two parameters in the analysis since the focus of this paper is non-scale free capabilities. See MacDonald and Ryall (2004: 1330-1331) and Adner and Zemsky (2008) for analytical examinations of scale-free capabilities and the effect of their fungibility on profitability.

firm will diversify first while the less capable firm stays focused on the initial market. Building on this property, we conduct the following numerical analysis to examine how the demand threshold to diversify changes with the degree of capability asymmetry. In so doing, we identify some interesting competitive effects of asymmetry in firms' capabilities.

Insert Figure 1 about here

In Figure 1, $\bar{a}_N$ is the demand threshold where the more capable firm (with capabilities T_1) first chooses to diversify, while the less capable firm (with capabilities T_2) is still focused. The downward movement along a given curve captures the "capacity effect," meaning that as the focal firm's own capabilities (T_1) increase, its threshold to diversify becomes smaller, or firm 1 reaches the threshold to diversify earlier. The shift in curves from square ($T_2 = 1.5$) to diamond ($T_2 = 0.5$) captures the "competition effect," meaning that as the competitor's capabilities (T_2) decrease, the threshold for the more capable firm (with capabilities T_1) to start diversifying becomes greater. This means that the more capable firm diversifies later (for a larger size of the new market), since the current market becomes more attractive.

Implications of numerical investigation for profit margin and diversification

Having characterized diversification behavior under Cournot competition, we now address the central results regarding the impact of diversification on profit margins. In particular, the numerical analysis examines whether Proposition 1 regarding the tradeoff between total profits and profit margins holds under Cournot competition.

In Proposition 1, under the Bertrand setting, we compare the profit margin of a diversified firm with a hypothetical benchmark of the weighted average of the firm focusing

separately on the two markets. However, in the Bertrand setting, since there is essentially a single firm, we cannot make cross-sectional comparisons of firms with heterogeneous capabilities. The following analysis in the Cournot setting addresses this issue. With the same baseline parameter values of $a_I = 10$, $b_I = b_N = 1, 1-\delta = 0.8$, we let the more capable firm have capabilities $T_1 = 2$ and the less capable one have capabilities $T_2 = 1.5$.

Insert Figure 2 about here

Examining profit margins as the size of the new market varies we see a number of interesting properties (Figure 2). First, consistent with Proposition 1, for both the more and less capable firm, there is a substantial drop in profit margin when the relative demand level in the new market passes the threshold sufficient to elicit diversification and market entry. Second, consistent with Proposition 2, given the significant heterogeneity in capabilities the first firm to diversify is the more capable firm. In the interim range of market demand for which it is optimal for the high-capability firm to diversify but for the low-capability firm to remain focused on the initial market, we see that the more capable firm may end up with a lower profit margin than the less capable firm.[12] A firm's rational diversification decision is driven by its pursuit of profit maximization rather than profit margin maximization. Thus, profit margin may not be a good proxy indicator for differences in capabilities among firms that vary in their degree of diversification.

Another finding is that while both firms' profit margins drop substantially when the relative demand level in the new market passes the threshold sufficient to elicit diversification, their profit margins will gradually increase as the relative demand grows further. Moreover,

[12] Two working papers show a similar result by assuming multiple equal-sized markets that are perfectly competitive (Santalo, 2002) or monopolistically competitive (Nocke and Yeaple, 2008).

when the demand level in the new market reaches a level sufficient for the low-capability firm to diversify as well, we see the cross-sectional property that one generally intuits that the more capable firm earns a higher profit margin.

Finally, note that if the capability differences among the two firms is rather extreme, it may be the case that the more capable firm maintains a higher profit margin for all demand environments, even in that intermediate level of new market demand that generates diversification behavior on the part of the more capable firm but a focus on the original market on the part of the less capable firm. Figure 3 provides an illustration of this with: $a_I = 10$, $b_I = b_N = 1$, $1-\delta = 0.8$, $T_1 = 2$, $T_2 = 0.5$.

Insert Figure 3 about here

DISCUSSION: IMPLICATIONS OF OPPORTUNITY COSTS AND DIVERSIFICATION

While the contemporary literature on diversification from a resource-based view builds upon the idea of excess firm capabilities developed in Penrose (1959), the emphasis has been on the fungibility of resources across domains. Making a clear analytical distinction between scale-free capabilities and those that are subject to opportunity costs and must be allocated to one use or another helps to shift the discourse back to Penrose's (1959) focus on the stock of organizational capabilities. The existence of non-scale free capabilities implies that profit-maximizing diversification decisions should be based on 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 capabilities diminishes across markets. We further identify the demand thresholds for firms to diversify as a function of their capabilities, which

allows us to infer the effect of heterogeneous capabilities on the order of diversification in the face of competitive interactions. The recognition of capabilities that must be allocated across multiple segments based on opportunity costs also provides a profit-maximizing explanation for the divergence between total profits and profit margins and in turn an alternative explanation of the diversification discount.

Our developed model suggests an alternative self-selection mechanism that can account for the observation of a cross-sectional diversification discount (Berger and Ofek, 1995; Lang and Stulz, 1994). Firms with superior capabilities in a low-value existing market context enhance their profits by diversifying, but at the same time incur lower average return due to the spread of non-scale free capabilities across applications. Therefore, it may not be, as suggested by Gomes and Livdan (2004), that those firms with fewer capabilities (lower 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 diversifying 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 non-scale free capabilities in less favorable opportunities. Our argument suggests that diversifying firms are "good types" (i.e., high capabilities) operating in "bad" market contexts.

This argument suggests that 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 the overall market maturity. In either case, such profit-maximizing 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 non-scale free 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 fine-grained industry classification is able to offer more refined measures of firms' skills and capabilities than cross-industry analysis that inevitably must rely on more coarse-grained data. Thus, the analysis that contrasts *de novo* entrants versus *de alia* entrants, such as Klepper and Simons (2000), Carroll *et al*. (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 that firms face. Even industry classification at the four-digit SIC level may incorporate many rather distinct submarkets with quite different demand patterns.[13] 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 non-scale free 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 correspondingly relatively *ex-ante* weak performance as suggested by recent writings in the corporate finance literature (e.g., Gomes and Livdan, 2004)?

[13] As an illustration of the heterogeneity within a four-digit SIC class, consider the cardiovascular medical device industry. Although this industry is mainly underneath primary SIC 3841 (surgical and medical instruments and apparatus) and 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 eight independent product sub-markets, such as stents, pacemakers, and heart valves, that have experienced very different industry life cycles (Wu, 2009).

Identifying the crucial role of opportunity cost of resource allocation provides an alternative, economic-based explanation for the reluctance of established firms to aggressively enter new domains as identified by Christensen (1997) in the context of his work on the disk industry. The established firms in the disk drive industry faced a critical choice as to how to allocate their valuable product development teams in the face of brutal competition with rapid model introductions in their existing technological platforms or to introduce drives with an alternative format. Based on the relative size of the market for their current family of drives, as compared to the emerging market for smaller drives and ex-ante assessments of the growth rates of these alternative technologies, straightforward opportunity cost logic could argue for the apparent inertia of the established firms.

The notion of opportunity cost is a powerful concept. Perhaps its basic and pervasive nature may cause us to ignore, or at least under-play, its role. 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. Ultimately, we need to develop explicit characterizations of these "supply side" dynamics of firms' capabilities in conjunction with analyses of the competitive dynamics of product market competition. Both the investment in capabilities, as well as their allocation across contexts, is a function of firms' perception of their demand environments and the competitive conditions they face. Recent work has rightly highlighted the endogeneity of diversification decisions. However, in capturing the basis of this endogeneity, it is important to leverage our existing insights regarding corporate diversification. The economic logic of the allocation of non-scale free capabilities over alternative domains builds on traditional capability-based views of the firm and provides a basis for endogenous diversification decisions consistent with existing and emerging empirical findings.

REFERENCES

Aaker DA. 2004. *Brand Portfolio Strategy: Creating Relevance, Differentiation, Energy, Leverage, and Clarity*. Free Press: New York, NY.

Adner R, Zemsky P. 2008. Diversification and performance: Linking relatedness, market structure, relatedness and the decision to diversify. Working Paper, INSEAD.

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

Berger PG, Ofek E. 1995. Diversifications effect on firm value. *Journal of Financial Economics* **37**(1): 39-65.

Bettis RA. 1981. Performance differences in related and unrelated diversified firms. *Strategic Management Journal* **2**(4): 379-393.

Burgelman RA. 1994. Fading memories: A process theory of strategic business exit in dynamic environments. *Administrative Science Quarterly* **39**(1): 24-56.

Campa JM, Kedia S. 2002. Explaining the diversification discount. *Journal of Finance* **57**(4): 1731-1762.

Carroll GR, Bigelow LS, Seidel M-DL, Tsai LB. 1996. The fates of de novo and de alio producers in the American automobile industry 1885-1981. *Strategic Management Journal* **17**(Evolutionary Perspectives on Strategy Supplement): 117-137.

Chandler AD, Jr. 1969. The structure of American industry in the twentieth century: A historical overview. *The Business History Review* **43**(3): 255-298.

Chandler ADJ. 1977. *The Visible Hand: The Managerial Revolution in American Business*. Belknap Press: Cambridge, MA.

Chang SJ. 1995. International expansion strategy of Japanese firms: Capability building through sequential entry. *The Academy of Management Journal* **38**(2): 383-407.

Christensen C. 1997. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press: Boston, MA.

Christensen CM, Bower JL. 1996. Customer power, strategic investment, and the failure of leading firms. *Strategic Management Journal* **17**(3): 197-218.

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

Hannan MT, Freeman J. 1989. *Organizational Ecology*. Harvard University Press: Cambridge, MA.

Helfat CE. 2003. Stylized facts regarding the evolution of organizational resources and capabilities. In CE Helfat (Ed.), *The Blackwell/Strategic Management Society Handbook of Organizational Capabilities: Emergence, Development, and Change*: 1-11. Blackwell Publishers: Malden, MA.

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

Helfat CE, Lieberman MB. 2002. The birth of capabilities: Market entry and the importance of pre-history. *Industrial & Corporate Change* **11**(4): 725-760.

Hitt MA, Hoskisson RE, Ireland RD, Harrison JS. 1991. Effects of acquisitions on R&D inputs and outputs. *The Academy of Management Journal* **34**(3): 693-706.

Jensen MC. 1986. Agency costs of free cash flow, corporate-finance, and takeovers. *American Economic Review* **76**(2): 323-329.

Klepper S. 1996. Entry, exit, growth, and innovation over the product life cycle. *American Economic Review* **86**(3): 562-583.

Klepper S, Simons KL. 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.

Kreps DM, Scheinkman JA. 1983. Quantity precommitment and Bertrand competition yield Cournot outcomes. *The Bell Journal of Economics* **14**(2): 326-337.

Lang LHP, Stulz RM. 1994. Tobin's q, corporate diversification, and firm performance. *Journal of Political Economy* **102**(6): 1248-1280.

Lindenberg EB, Ross SA. 1981. Tobin's *q* ratio and industrial organization. *The Journal of Business* **54**(1): 1-32.

Lippman S, Rumelt R. 1982. Uncertain imitability: An analysis of interfirm differences in efficiency under competition. *Bell Journal of Economics* **13**(2): 418-453.

MacDonald G, Ryall MD. 2004. How do value creation and competition determine whether a firm appropriates value? *Management Science* **50**(10): 1319-1333.

Mahoney JT, Pandian JR. 1992. The resource-based view within the conversation of strategic management. *Strategic Management Journal* **13**(5): 363-380.

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

Markides CC, Williamson PJ. 1994. Related diversification, core competences and corporate performance. *Strategic Management Journal* **15**(Special Issue): 149-165.

Mitchell W. 1989. Whether and when? Probability and timing of incumbents' entry into emerging industrial subfields. *Administrative Science Quarterly* **34**(2): 208-230.

Montgomery CA, Wernerfelt B. 1988. Diversification, Ricardian rents, and Tobin's q. *Rand Journal of Economics* **19**(4): 623-632.

Nocke V, Yeaple S. 2008. Globalization and the size distribution of multiproduct firms. Working Paper, Oxford University

Palich LE, Cardinal LB, Miller CC. 2000. Curvilinearity in the diversification-performance linkage: An examination of over three decades of research. *Strategic Management Journal* **21**: 155-174.

Penrose ET. 1959. *The Theory of the Growth of the Firm*. Blackwell Publishers: Oxford.

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

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

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

Rubin PH. 1973. The expansion of firms. *Journal of Political Economy* **81**(4): 936-949.

Rumelt RP. 1974. *Strategy, Structure and Economic Performance*. Harvard Business School Press: Boston MA.

Santalo J. 2002. Organizational capital and the existence of a diversification and size discount Working Paper, University of Chicago

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 DJ. 1980. Economies of scope and the scope of the enterprise. *Journal of Economic Behavior and Organization* **1**(3): 223-247.

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

Uzawa H. 1969. Time preference and the Penrose effect in a two-class model of economic growth. *The Journal of Political Economy* **77**(4): 628-652.

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

Winter S. 1987. Knowledge and competence as strategic assets. In DJ Teece (Ed.), *The Competitive Challenge: Strategies for Industrial Innovation and Renewal*: 159-184. Ballinger Pub. Co.: Cambridge, MA.

Winter SG. 1995. The four Rs of profitability: Rents, resources, routines, and replication. In CA Montgomery (Ed.), *Resource-based and Evolutionary Theories of the Firm: Towards a Synthesis*: 147-178. Kluwer Academic Publishers: Boston.

Winter SG, Szulanski G. 2001. Replication as strategy. *Organization Science* **12**(6): 730-743.

Wu B. 2009. Opportunity costs, relative demand maturity, and corporate diversification: Evidence from the cardiovascular medical device industry, 1976-2004. Mack Center Working Paper, Wharton School.

Table 1: Dimensions of capabilities

	Non-scale-free (positive opportunity cost)	Scale-free (zero opportunity cost)
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

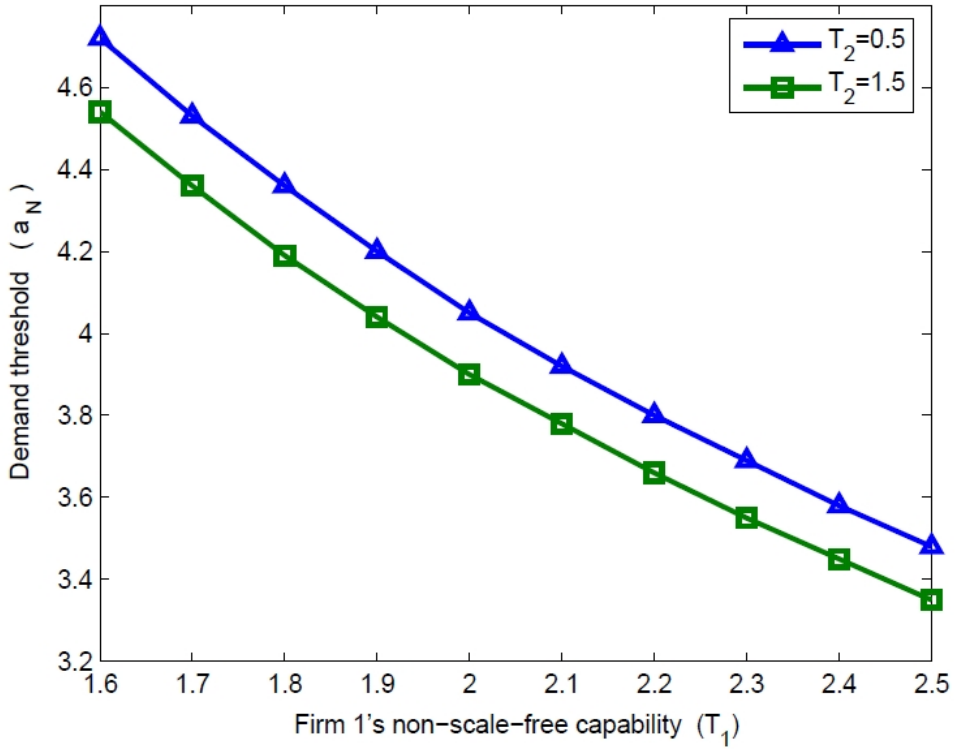


Figure 1: Capability asymmetry and demand threshold

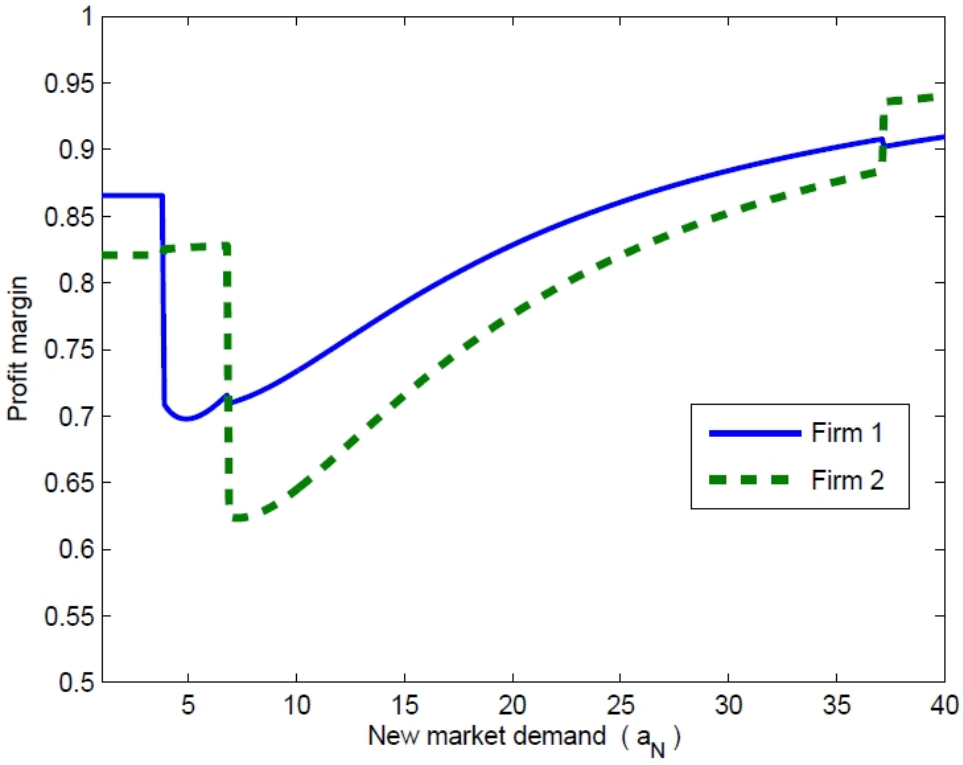


Figure 2: Change in profit margin along with relative demand

$(a_I = 10,\ b_I = b_N = 1, T_1 = 2,\ T_2 = 1.5,\ \text{and}\ (1-\delta) = 0.8)$

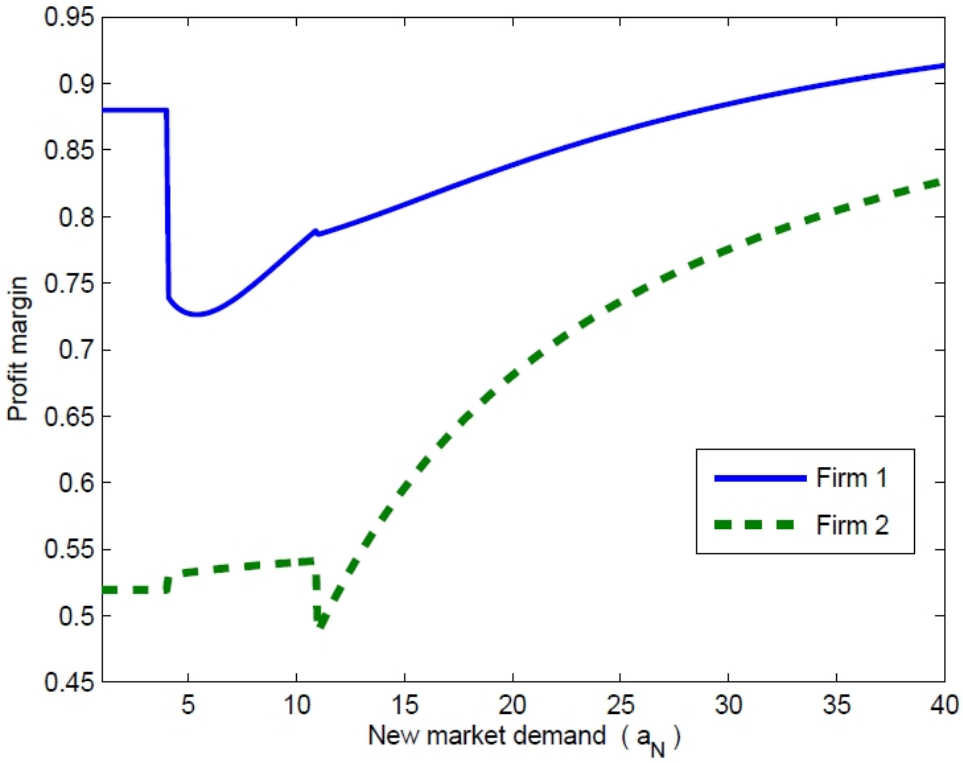


Figure 3: Change in profit margin along with relative demand

$(a_I = 10,\ b_I = b_N = 1, T_1 = 2,\ T_2 = 0.5,\ \text{and}\ (1-\delta) = 0.8)$

APPENDIX 1

In proving Proposition 1, we first solve the interior solution (t_I^*, t_N^*) to the maximization problem (2) in the text when demand conditions are such that it is optimal for the firm to diversify:[14]

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

Inserting (t_I^*, t_N^*) into the profit margin of the diversification strategy π^* in equation (3) in the text, we can transform equation (3) as

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

Notice that the first term (in the square brackets) in the above equation is exactly the weighted average of two focus strategies in equation (4) in the text. Therefore, equation (5) in the text and thus Proposition 1 is proved. ■

[14] It should be noted that while competitors' cost efficiency does not affect the continuous allocation of non-scale-free capabilities to a certain segment in the Bertrand setting, it does influence the boundary conditions that determine whether a firm chooses to enter a segment.

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 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 is $k_I = \frac{s_I}{\gamma_I T}$; (ii) if all capabilities are focused in the new market, then capital 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 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:

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

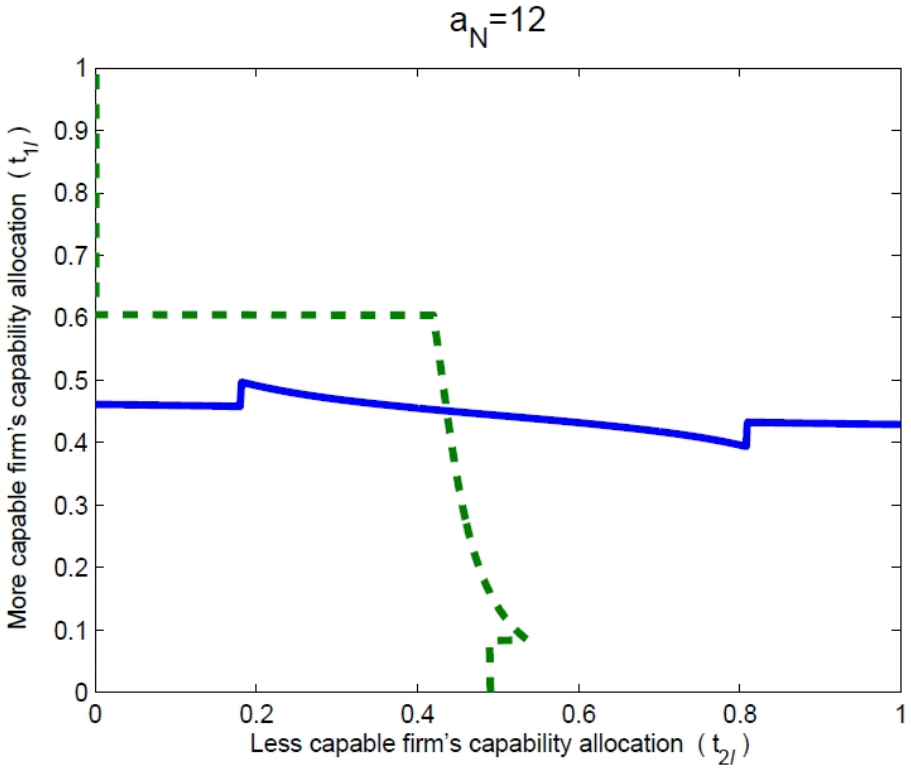
Finally, we compare the Tobin's q values associated with these two strategies by examining their ratio:

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

Therefore, there exists a discount factor

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

APPENDIX 3



The above figure indicates the best response curves when relative demand conditions are such that both firms diversify in equilibrium. The solid line illustrates the reasons why firm 1's best response curve is neither continuous nor monotonic. Firm 1's best response curve has small jumps both when t_{2I} is around 0.2 and 0.8. The "jump" around the value of $t_{2I} = 0.2$ stems from the fact that when $t_{2I} < 0.2$, firm 2's cost value is sufficiently high in the initial market that it is not able to produce in the initial market. As a result, firm 1's best response allocation of capabilities has a jump up as t_{2I} increases and passes 0.2, because firm 2 begins to produce in the initial market. Similarly, when $t_{2I} > 0.82$, firm 2's cost is too high in the new market to be viable and produce a non-zero quantity. As a result, firm 1's best response allocation of capabilities jumps up as t_{2I} increases and passes 0.82, because at that point firm 2 exits participation in the new market and, as a result, firm 1 does not need to put so many capabilities into the new market.

Similarly, the dotted line illustrates the reasons why firm 2's best response curve is neither continuous nor monotonic. When $t_{1I} = 0.6$, firm 2's best response curve jumps from 0 to 0.42. This jump stems from firm 2's entry into the new market. In addition, Firm 2's best response curve has a small jump when $t_{1I} = 0.1$. This "jump" stems from the fact that when $t_{1I} < 0.1$, firm 1's cost value is sufficiently high in the initial market that it is not able to produce in the initial market. As a result, firm 2's best response allocation has a jump up as t_{1I} increases and passes 0.1, because firm 1 begins to produce in the initial market.

APPENDIX 4

Note that to capture the change in relative demand, without loss of generality, for all the analyses, we hold constant the size of the initial market a_I and only vary the size of the new market a_N. Let $P_1 \triangleq P(a_N, T_1, T_2, t_{1I}, t_{2I})$ be firm 1's profit function given the new market size a_N, firm 1's capability T_1, the competitor firm 2's capability T_2, firm 1's allocation t_{1I}, and the competitor firm 2's allocation t_{2I}. Similarly, $P_2 \triangleq P(a_N, T_2, T_1, t_{2I}, t_{1I})$ represents firm 2's profit function given the new market size a_N, firm 2's capability T_2, the competitor firm 1's capability T_1, firm 2's allocation t_{2I}, and the competitor firm 1's allocation t_{1I}. Note that $P(a_N, T_1, T_2, t_{1I}, t_{2I}) = \pi_{1I} + \pi_{1N}$ and $P(a_N, T_2, T_1, t_{2I}, t_{1I}) = \pi_{2I} + \pi_{2N}$, where,

$$\pi_{im} = b_m (q_{im})^2 \text{ and } q_{im} = \begin{cases} \dfrac{a_m + c_{(3-i)m} - 2c_{im}}{3b_m} & , if\ q_{(3-i)m} > 0 \\ \dfrac{a_m - c_{im}}{2b_m} & , if\ q_{(3-i)m} = 0 \end{cases} \quad \text{for } i = 1,2;\ m = I, N \qquad \text{(A1)}$$

Below we restate Proposition 2a by characterizing the properties of the best response curves. Moreover, without loss of generality, we fix the capability level of firm 1 and vary the capability level of firm 2 ($T_2 = T$).

Proposition 2a: Let $T_2 = T$. There exists an $\hat{a}_N > 0, \eta > 0, \text{and } \varepsilon > 0$, such that $P(a_N, T_1, T, t_{1I}, 1)$ reaches maximum at some $t_{1I} = t(a_N, T_1) < 1$ and $P(a_N, T, T_1, t_{2I}, t(a_N, T_1))$ has a unique maximum at $t_{2I} = 1$, whenever $\hat{a}_N < a_N < \hat{a}_N + \eta$ and $|T_1 - T| < \varepsilon$.

We start by proving when a_N is small enough, there is a unique equilibrium ($t_{1I} = 1, t_{2I} = 1$), or both firms focus on the initial market.

Lemma 1: Given any T_1 and T_2, there exists an $\underline{a}_N$ such that when $a_N < \underline{a}_N$, P_1 has a unique maximum at $t_{1I} = 1$ for all t_{2I} and P_2 has a unique maximum at $t_{2I} = 1$ for all t_{1I}.

Proof: We provide the proof for P_1, and that for P_2 is analogous. Note that $q_{1N} = 0$ and thus $\pi_{1N} = b_N (q_{1N})^2 = 0$ for all t_{2I} whenever $t_{1I} \geq 1 - \dfrac{1}{a_N (1-\delta)\gamma T_1}$ because $c_{1N} = \dfrac{1}{(1-\delta)\gamma(1-t_{1I})T_1} \geq a_N \geq p_N$; in other words, firm 1's marginal cost in the new market is always greater than the maximal possible price unless firm 1 allocates enough capability to the new market (i.e., $1 - t_{1I} \geq \dfrac{1}{a_N (1-\delta)\gamma T_1}$). This result together with the fact that q_{1I} is strictly

increasing in t_{1I} on $[1-\frac{1}{a_N(1-\delta)\gamma T_1},1]$ implies that $P_1=\pi_{1I}+\pi_{1N}=\pi_{1I}+0=b_I(q_{1I})^2$ is strictly increasing in t_{1I} on $[1-\frac{1}{a_N(1-\delta)\gamma T_1},1]$.

Note that when c we have $q_{1N}=0$ and hence that $P_1=b_I(q_{1I})^2$ is increasing in t_{1I} on [0, 1] (since q_{1I} is increasing in t_{1I}). Thus, when $a_N=0$, P_1 has a unique maximum at $t_{1I}=1$ for all t_{2I}. This result implies that there exists an $\underline{a}_N$ such that when $a_N<\underline{a}_N$ P_1 has a unique maximum at $t_{1I}=1$ for all t_{2I}, provided that P_1 is continuously increasing in a_N for $t_{1I}\in[0,1-\frac{1}{a_N(1-\delta)\gamma T_1}]$ and is constant with respect to the changes of a_N for $t_{1I}\in[1-\frac{1}{a_N(1-\delta)\gamma T_1},1]$ (this later condition implies that P_1 is still strictly increasing in t_{1I} on $[1-\frac{1}{a_N(1-\delta)\gamma T_1},1]$, as proved in the previous paragraph). ■

Lemma 1 immediately implies that when a_N is small enough, $(t_{1I},t_{2I})=(1,1)$, or both firms focusing on the initial market, is the unique equilibrium.

We then prove that when the new market gets sufficiently large and two firms' capability asymmetry is small enough, either firm may first diversify while the other stays focused at the initial market.

In so doing, we first prove a lemma to show that if it is optimal for firm 1 to choose $t_{1I}=1$ given some t_{2I}, it must be still optimal for firm 1 to choose $t_{1I}=1$ given some $\tilde{t}_{2I}<t_{2I}$. That is,

Lemma 2: $\frac{\partial^2 P_1}{\partial t_{2I}\partial t_{1I}}=2b_I\frac{\partial q_{1I}}{\partial t_{2I}}\frac{\partial q_{1I}}{\partial t_{1I}}+2b_N\frac{\partial q_{1N}}{\partial t_{2I}}\frac{\partial q_{1N}}{\partial t_{1I}}<0$.

Proof: $\frac{\partial P_1}{\partial t_{1I}}=2b_I q_{1I}\frac{\partial q_{1I}}{\partial t_{1I}}+2b_N q_{1N}\frac{\partial q_{1N}}{\partial t_{1I}}$,

$$\frac{\partial^2 P_1}{\partial t_{1I}\partial t_{2I}}=2b_I\frac{\partial q_{1I}}{\partial t_{1I}}\frac{\partial q_{1I}}{\partial t_{2I}}+2b_I q_{1I}\frac{\partial^2 q_{1I}}{\partial t_{1I}\partial t_{2I}}+2b_N\frac{\partial q_{1N}}{\partial t_{1I}}\frac{\partial q_{1N}}{\partial t_{2I}}+2b_N q_{1N}\frac{\partial^2 q_{1N}}{\partial t_{1I}\partial t_{2I}}$$

Lemma 2 follows because $\frac{\partial^2 q_{1I}}{\partial t_{2I}\partial t_{1I}}=0$ and $\frac{\partial^2 q_{1N}}{\partial t_{2I}\partial t_{1I}}=0$; $\frac{\partial q_{1I}}{\partial t_{2I}}\frac{\partial q_{1I}}{\partial t_{1I}}<0$ and $\frac{\partial q_{1N}}{\partial t_{2I}}\frac{\partial q_{1N}}{\partial t_{1I}}<0$ (Easy to check from equation A1). ■

Therefore, if it is optimal to focus on the initial market when the competitor focuses on the initial market ($t_{2I}=1$), it must be still optimal to focus if the competitor diversifies into the new market. Similarly this lemma applies to firm 2.

Lemma 3: Given $T>0$, when the two firms have the same level of capabilities, that is $T_1=T_2=T$, there exists an $\hat{a}_N>0$ such that $P(\hat{a}_N,T,T,t_{1I},1)$ has at least two maxima including one at $t_{1I}=1$. $P(a_N,T,T,t_{1I},1)$ has a unique maximum at $t_{1I}=1$ whenever $a_N<\hat{a}_N$, and $P(a_N,T,T,t_{1I},1)$ reaches maximum at some $t_{1I}=t_{1I}(a_N)<1$ whenever $a_N>\hat{a}_N$.

Proof: According to the argument in Lemma 1, it is easy to check that $P(0,T,T,t_{1I},1)$ is increasing in t_{1I} over [0,1] and is strictly increasing in t_{1I} when t_{1I} is close to 1. Thus, when a_N is small enough, $P(a_N,T,T,t_{1I},1)$ has a unique maximum at $t_{1I}=1$. Also, it is straightforward to check that $P(a_N,T,T,t_{1I},1)>P(a_N,T,T,1,1)$ for any $t_{1I}<1$ when a_N is big enough. This result together with the fact that $P(a_N,T,T,t_{1I},1)$ is continuous in a_N implies that there exists a threshold $\hat{a}_N>0$ such that $P(\hat{a}_N,T,T,t_{1I},1)$ has at least two maxima including one at $t_{1I}=1$.

Note that: $\dfrac{\partial^2}{\partial a_N \partial t_{1I}}P(a_N,T,T,t_{1I},1)<0$, which implies that $P(a_N,T,T,t_{1I},1)$ has a unique maximum at $t_{1I}=1$ whenever $a_N<\hat{a}_N$ and $P(a_N,T,T,t_{1I},1)$ reaches maximum at some $t_{1I}=t_{1I}(a_N)<1$ whenever $a_N>\hat{a}_N$. ■

Lemma 2 and Lemma 3 together imply the following corollary for all $t_{2I}<1$.

Corollary 1: $P(\hat{a}_N,T,T,t_{1I},t_{2I})$ has a unique maximum at $t_{1I}=1$ for all $t_{2I}<1$.

Now we can prove Proposition 2(a):

Proof: Proposition 2a follows from Lemma 1, Lemma 3, and Corollary 1 and the fact that P is differentiable with respect to all arguments. ■

Note that $P(a_N,T_1,T,t_{1I},1)$ reaches maximum at some $t_{1I}=t(a_N,T_1)<1$. This maximum is not necessarily unique. There can be a unique maximum at $t_{1I}=0$, or a unique maximum at $0<t_{1I}<1$, or two maxima at both $t_{1I}=0$ and some $0<t_{1I}<1$. However, in any case, Proposition 2a is proved in that, in equilibrium, either firm can diversify first (diversifies partly into the new market or completely switches to the new market) while the other firm stays focused in the initial market.

Below we restate Proposition 2b by characterizing the properties of the best response curves. Moreover, without loss of generality, we let firm 2 be the less capable firm.

Proposition 2b: Fixing T_2, there exists a $\tilde{T} > 0$, $\hat{a}_N > 0$, $\breve{a}_N > 0$, $\hat{a}_N < \breve{a}_N$, and $\eta > 0$ such that, when $T_1 > \tilde{T}$,

i) $P(a_N, T_1, T_2, t_{1I}, 1)$ has a unique maximum at $t_{1I} = 1$ for all $a_N < \hat{a}_N$

ii) $P(a_N, T_1, T_2, t_{1I}, 1)$ has a unique maximum at some $0 < t_{1I} < 1$ for all $a_N \in (\hat{a}_N, \hat{a}_N + \eta)$, and

iii) $P(a_N, T_2, T_1, t_{2I}, 1)$ has a unique maximum at $t_{2I} = 1$ for all $a_N < \breve{a}_N$

Proof: It is straightforward to see that, for firm 1, there exists an $\hat{a}_N > 0$ such that $P(\hat{a}_N, T_1, T_2, t_{1I}, 1)$ has a maximum at $t_{1I} = 1$ and a maximum at some $\bar{t}_{1I} < 1$. Similar to the proof of Lemma 3, $\frac{\partial^2}{\partial a_N \partial t_{1I}} P(a_N, T_1, T_2, t_{1I}, 1) < 0$, so $P(a_N, T_1, T_2, t_{1I}, 1)$ has a unique maximum at $t_{1I} = 1$ for all $a_N < \hat{a}_N$, and $P(a_N, T_1, T_2, t_{1I}, 1)$ reaches maximum at some $t_{1I} < 1$ for all $a_N \in (\hat{a}_N, \hat{a}_N + \eta)$ (not necessarily unique, since there can be two maxima at both $t_{1I} = 0$ and some $0 < t_{1I} < 1$).

Moreover, $\hat{a}_N > 0$ satisfies $P(\hat{a}_N, T_1, T_2, \bar{t}_{1I}, 1) = P(\hat{a}_N, T_1, T_2, 1, 1)$. That is,

$$\frac{(a_I + \frac{1}{\gamma T_2} - \frac{2}{\gamma \bar{t}_{1I} T_1})^2}{9b_I} + \frac{(\hat{a}_N - \frac{1}{(1-\delta)\gamma(1-\bar{t}_{1I})T_1})^2}{4b_N} = \frac{(a_I + \frac{1}{\gamma T_2} - \frac{2}{\gamma T_1})^2}{9b_I}.$$ It is easy to check that $\hat{a}_N$ approaches zero as T_1 increases. In addition, as T_1 increases, $P(\hat{a}_N, T_1, T_2, t_{1I}, 1)$ has a unique maximum at some $0 < \bar{t}_{1I} < 1$, since $\bar{t}_{1I} = 0$ can never be a maximum when $\hat{a}_N$ approaches zero. This fact implies that $P(a_N, T_1, T_2, t_{1I}, 1)$ has a unique maximum at some $0 < t_{1I} < 1$ for all $a_N \in (\hat{a}_N, \hat{a}_N + \eta)$.

Similarly, for firm 2, it is straightforward to see that there exists an $\breve{a}_N > 0$ such that $P(\breve{a}_N, T_2, T_1, t_{2I}, 1)$ has a maximum at $t_{2I} = 1$ and a maximum at some $\bar{t}_{2I} < 1$.
Namely, $P(\breve{a}_N, T_2, T_1, \bar{t}_2, 1) = P(\breve{a}_N, T_2, T_1, 1, 1)$. That is:

$$\frac{(a_I + \frac{1}{T_1} - \frac{2}{\gamma \bar{t}_{2I} T_2})^2}{9b_I} + \frac{(\breve{a}_N - \frac{1}{(1-\delta)\gamma(1-\bar{t}_{2I})T_2})^2}{4b_N} = \frac{(a_I + \frac{1}{\gamma T_1} - \frac{2}{\gamma T_2})^2}{9b_I}.$$

Since $\frac{\partial^2}{\partial a_N \partial t_{2I}} P(a_N, T_2, T_1, t_{2I}, 1) < 0$, $P(a_N, T_2, T_1, t_{2I}, 1)$ has a unique maximum at $t_{2I} = 1$ for all $a_N < \breve{a}_N$. In addition, it is straightforward to check that $\breve{a}_N$ approaches a constant as T_1 increases. Since $\hat{a}_N$ approaches zero as T_1 increases, $\hat{a}_N < \breve{a}_N$ when T_1 is large enough. Since P_1 is

continuously increasing in T_1 and P_2 is continuously decreasing in T_1, there exists a $\tilde{T} > 0$; for all $\hat{a}_N < a_N < \hat{a}_N + \eta < \breve{a}_N$, it is a unique equilibrium that $0 < t_{1I} < 1$ and $t_{2I} = 1$ when $T_1 > \tilde{T}$.■

Below we restate Proposition 2c by characterizing the properties of the best response curves. Moreover, without loss of generality, we let firm 2 be the less capable firm.

Proposition 2c: Given a_I and T_1, there exists $\underline{T} > 0$ such that for any $0 < T_2 < \underline{T}$ and $0 \le t_{1I} \le 1$, $P(a_N, T_1, T_2, t_{1I}, t_{2I})$ achieves maximum at $t_{2I} = 0$ or $t_{2I} = 1$.

Proof: Recall that the total profit of firm 2 is $P(a_N, T_2, T_1, t_{2I}, t_{1I}) = \pi_{2I} + \pi_{2N}$, where

$$\pi_{2m} = b_m (q_{2m})^2 \text{ and } q_{2m} = \begin{cases} \max(\frac{a_m + c_{1m} - 2c_{2m}}{3b_m}, 0), & \text{if } q_{1m} > 0 \\ \max(\frac{a_m - c_{2m}}{2b_m}, 0), & \text{if } q_{1m} = 0 \end{cases} \quad \text{for } m = I, N$$

Given that $c_{1I} = \frac{1}{\gamma t_{1I} T_1}$, $c_{2I} = \frac{1}{\gamma t_{2I} T_2}$, $c_{1N} = \frac{1}{(1-\delta)\gamma(1-t_{1I})T_1}$, and $c_{2N} = \frac{1}{(1-\delta)\gamma(1-t_{2I})T_2}$, it is easy to check the following are true:

1) q_{2I} (and hence π_{2I}) is increasing in t_{2I} and T_2. There exists $\underline{t}_{2I}^{I} \in [0,1]$ such that $q_{2I} = 0$ (and hence $\pi_{2I} = 0$) for all $t_{2I} \in [0, \underline{t}_{2I}^{I})$. The threshold $\underline{t}_{2I}^{I}$ decreases in $_{T_2}$ and equals 1 when T_2 is small enough. $\underline{t}_{2I}^{I}$ does not change with a_N.

2) q_{2N} (and hence π_{2N}) is decreasing in t_{2I}, but increasing in T_2. There exists $\underline{t}_{2I}^{N} \in [0,1]$ such that $q_{2N} = 0$ (and hence $\pi_{2N} = 0$) for all $t_{2I} \in (\underline{t}_{2I}^{N}, 1]$. The threshold $\underline{t}_{2I}^{N}$ increases in a_N and approaches 1 as a_N approaches infinity. $\underline{t}_{2I}^{N}$ is increasing in T_2.

Therefore, there exists an $\tilde{a}_N > 0$ and a threshold $\underline{T}(\tilde{a}_N) > 0$ such that $\underline{t}_{2I}^{N} \in (0,1)$ and whenever $0 < T_2 < \underline{T}(\tilde{a}_N)$ we have $\underline{t}_{2I}^{I} > \underline{t}_{2I}^{N}$ (since $\underline{t}_{2I}^{I}$ is decreasing in T_2 and $\underline{t}_{2I}^{N}$ is increasing in T_2 as explained above) and $\pi_{2I}(t_{2I} = 1) < \pi_{2N}(t_{2I} = 0)$ (since $\pi_{2I}(t_{2I} = 1)$ is independent of a_N while $\pi_{2N}(t_{2I} = 0)$ is increasing in a_N as explained above). This fact implies that given any a_I, T_1, and t_{1I}, firm 2's total profit $P(a_N, T_2, T_1, t_{2I}, t_{1I}) = \pi_{2I} + \pi_{2N}$ achieves maximum at $t_{2I} = 0$ or $t_{2I} = 1$ for all $a_N > 0$ whenever $T_2 < \underline{T}(\tilde{a}_N)$. This result clearly holds when $a_N = \tilde{a}_N$, since, by the above definition of $\tilde{a}_N$ and $\underline{T}(\tilde{a}_N)$, $P(a_N, T_2, T_1, t_{2I}, t_{1I}) = \pi_{2I} + \pi_{2N}$ achieves maximum at $t_{2I} = 0$. Next we prove this result for $a_N < \tilde{a}_N$ and $a_N > \tilde{a}_N$ respectively.

When $a_N < \tilde{a}_N$, we have $\underline{t}_{2I}^{I} > \underline{t}_{2I}^{N}$, since $\underline{t}_{2I}^{I}$ does not change with a_N and $\underline{t}_{2I}^{N}$ increases in a_N. Therefore, $P(a_N, T_2, T_1, t_{2I}, t_{1I}) = \pi_{2I} + \pi_{2N}$ achieves maximum at $t_{2I} = 0$ or $t_{2I} = 1$ because

$P(a_N, T_2, T_1, t_{2I}, t_{1I}) = \pi_{2I} + \pi_{2N}$ is strictly decreasing in t_{2I} on $[0, \underline{t}_{2I}^{N})$, is zero on $[\underline{t}_{2I}^{N}, \underline{t}_{2I}^{I}]$, and is strictly increasing in t_{2I} on $(\underline{t}_{2I}^{I}, 1]$.

When $a_N > \tilde{a}_N$, $P(a_N, T_2, T_1, 0, t_{1I}) = \pi_{2N}(a_N, t_{2I} = 0)$ is strictly greater than $P(a_N, T_2, T_1, t_{2I}, t_{1I}) = \pi_{2N}(a_N, t_{2I}) + \pi_{2I}(t_{2I})$ for any $t_{2I} \in (0,1]$, because $\pi_{2N}(a_N, t_{2I} = 0) - [\pi_{2N}(a_N, t_{2I}) + \pi_{2I}(t_{2I})]$ equals $\pi_{2N}(a_N, t_{2I} = 0) - \pi_{2N}(a_N, t_{2I}) > 0$ for all $t_{2I} \in (0, \underline{t}_{2I}^{I}]$, and is greater than $\pi_{2N}(a_N, t_{2I} = 0) - [\pi_{2N}(a_N, \underline{t}_{2I}^{I}) + \pi_{2I}(t_{2I} = 1)] > 0$ for $t_{2I} \in (\underline{t}_{2I}^{I}, 1]$ ($\pi_{2N}(a_N, t_{2I}) < \pi_{2N}(a_N, \underline{t}_{2I}^{I})$ for $t_{2I} \in (\underline{t}_{2I}^{I}, 1]$ since $\pi_{2N}(a_N, t_{2I})$ is decreasing in t_{2I} for $t_{2I} \in (\underline{t}_{2I}^{I}, 1]$; $\pi_{2I}(t_{2I}) \le \pi_{2I}(t_{2I} = 1)$ for $t_{2I} \in (\underline{t}_{2I}^{I}, 1]$ since $\pi_{2I}(t_{2I})$ is increasing in t_{2I} for $t_{2I} \in (\underline{t}_{2I}^{I}, 1]$).

Note that $\pi_{2N}(a_N, t_{2I} = 0) - [\pi_{2N}(a_N, \underline{t}_{2I}^{I}) + \pi_{2I}(t_{2I} = 1)] > 0$ is because $\pi_{2N}(a_N, t_{2I} = 0) - [\pi_{2N}(a_N, \underline{t}_{2I}^{I}) + \pi_{2I}(t_{2I} = 1)] >$ $\pi_{2N}(\tilde{a}_N, t_{2I} = 0) - [\pi_{2N}(\tilde{a}_N, \underline{t}_{2I}^{I}) + \pi_{2I}(t_{2I} = 1)] = \pi_{2N}(\tilde{a}_N, t_{2I} = 0) - [0 + \pi_{2I}(t_{2I} = 1)] > 0$, where the first inequality holds because $\frac{\partial^2}{\partial a_N \partial t_{2I}} P(a_N, T_2, T_1, t_{2I}, 1) < 0$ and $a_N > \tilde{a}_N$, the equality holds due to the definition of $\tilde{a}_N$ (when $a_N = \tilde{a}_N$, $\underline{t}_{2I}^{I} > \underline{t}_{2I}^{N}$ and $\pi_{2N}(a_N, t_{2I}) = 0$ for all $t_{2I} \in (\underline{t}_{2I}^{N}, 1]$), and the last inequality holds due to the definition of $\tilde{a}_N$ and $\underline{T}(\tilde{a}^n)$.

Therefore, $\underline{T}(\tilde{a}_N)$ serves as an upper bound such that, given a_I and T_1, the total profit of firm 2 achieves maximum at $t_{2I} = 0$ or $t_{2I} = 1$ for any $0 < T_2 < \underline{T}(\tilde{a}_N)$ and $0 \le t_{1I} \le 1$.■